

**MODELO DE PRONÓSTICO DE SERIES DE TIEMPO BASADO EN  
TÉCNICAS DE ANALÍTICA PREDICTIVA EN LA MEJORA DEL  
PROCESO DE DEFINICIÓN DEL PLAN Y PRESUPUESTO DE VENTAS**

**AUTOR**

**LUIS CARLOS SUAREZ BERNAL**

**ESCUELA COLOMBIANA DE INGENIERÍA  
FACULTAD DE INGENIERÍA DE SISTEMAS  
MAESTRIA EN GESTIÓN DE INFORMACIÓN  
BOGOTÁ**

**2018**

**MODELO DE PRONÓSTICO DE SERIES DE TIEMPO BASADO EN ii  
TÉCNICAS DE ANALÍTICA PREDICTIVA EN LA MEJORA DEL  
PROCESO DE DEFINICIÓN DEL PLAN Y PRESUPUESTO DE VENTAS**

**AUTOR**

**LUIS CARLOS SUAREZ BERNAL**

**Presentado para optar al título de: Magister en Gestión de Información**

**DIRECTOR**

**CARLOS RODRIGO RUIZ CRUZ**

**ESCUELA COLOMBIANA DE INGENIERÍA  
FACULTAD DE INGENIERÍA DE SISTEMAS  
MAESTRIA EN GESTIÓN DE INFORMACIÓN  
BOGOTÁ**

**2018**



## **Dedicatoria**

iv

Dedico este trabajo de grado a la inteligencia suprema y energía infinita que gobierna y da orden al universo.

## **Agradecimientos**

v

Gracias a mi familia por estar en los momentos cruciales en esta maravillosa experiencia.

La desarticulación de los procesos clave de la cadena de suministro en el proceso de planificación de la demanda, y específicamente en la definición del plan y del presupuesto de ventas y la toma de decisiones administrativas basadas en pronósticos de la demanda mediante la utilización de métodos causales (regresión lineal) y métodos cualitativos (de juicio), ejecutados de forma manual y de las decisiones tomadas con base en ellos, repercuten entre otros en una producción la cual no se ajustada a las necesidades reales de los clientes, a la baja optimización de inventarios, pérdida de oportunidades de ventas al no mantener un adecuado stock de inventario y en general a una deficiente programación de compras. Se ha planteado como objetivo definir el modelo más adecuado para el pronóstico de la demanda basado en técnicas de analítica predictiva, en la definición del plan y el presupuesto de ventas para la organización objeto de estudio. La metodología que se adoptará para la definición del modelo y de la aplicación de los resultados al entorno de negocio será CRISP-DM (Cross Industry Standard Process for Data Mining), la cual describe de forma normalizada el ciclo de vida de un proyecto estándar de análisis de datos. De acuerdo con la metodología, el ciclo de vida del proyecto consta de seis fases las cuales son: la comprensión del negocio, el estudio y comprensión de los datos, la preparación de los datos, el modelado (definición y calibración de modelos), la evaluación de los modelos y finalmente el despliegue del modelo con el fin de integrar el modelo con mejor desempeño al entorno de negocio, específicamente en el proceso de definición del plan y el presupuesto de ventas. Luego del análisis de la situación actual en la definición del problema, se realiza el planteamiento de la solución con base en los objetivos y a las prioridades definidas por el negocio. Se define como Fase I del proyecto, la automatización de la generación de pronósticos de ventas para la definición del presupuesto y consolidación del plan de ventas, como parte del proceso de planeación general de

la demanda en la organización, se define el alcance a alto nivel de las fases posteriores delvii proyecto y por ultimo se realiza la definición e implementación de la estrategia para la adopción del modelo, mediante el desarrollo de un sistema de información el cual permite en gran medida la automatización de proceso definidos.

**Palabras Clave:** pronóstico de la demanda, planificación de la demanda, plan de ventas, presupuesto de ventas, analítica predictiva, modelos cuantitativos, análisis series de tiempo.

The dismantling of the key processes of the supply chain in the process of demand planning, and specifically in the definition of the plan and the sales budget and the taking of administrative decisions based on forecasts of demand through the use of causal methods (linear regression) and qualitative methods (of judgment), executed manually and the decisions taken based on them, have repercussions among others in a production which is not adjusted to the real needs of the clients, to the low optimization of inventories, loss of sales opportunities due to not maintaining an adequate stock of inventory and, in general, poor purchasing scheduling. The objective was to define the most appropriate model for forecasting demand based on predictive analytical techniques, in the definition of the plan and the sales budget for the organization under study. The methodology that will be adopted for the definition of the model and the application of the results to the business environment will be CRISP-DM (Cross Industry Standard Process for Data Mining), which describes in a standardized way the life cycle of a standard project. analysis of data. According to the methodology, the life cycle of the project consists of six phases which are: the understanding of the business, the study and understanding of the data, the preparation of the data, the modeling (definition and calibration of models), the evaluation of the models and finally the deployment of the model in order to integrate the model with better performance to the business environment, specifically in the process of defining the plan and the sales budget. After the analysis of the current situation in the definition of the problem, the approach of the solution is made based on the objectives and priorities defined by the business. It is defined as Phase I of the project, the automation of the generation of sales forecasts for the definition of the budget and consolidation of the sales plan, as part of the process of general planning of the demand in the organization, the scope is defined at a high level of the later phases of the project and, finally, the

definition and implementation of the strategy for the adoption of the model, through the development of an information system which allows to a large extent the process automation defined.

**Key words:** demand forecast, demand planning, sales plan, sales budget, predictive analytics, quantitative models, time series analysis.

## Tabla de Contenidos

x

Capítulo 1 Introducción y consideraciones generales.....	1
Introducción.....	1
Consideraciones generales.....	2
Capítulo 2 Antecedentes.....	3
Planteamiento del problema.....	3
Enunciado del problema.....	3
Formulación del problema.....	7
Objetivos.....	8
Objetivo general.....	8
Objetivos específicos.....	8
Capítulo 3 Delimitación del alcance del proyecto.....	10
Proceso general de planeación de la demanda.....	11
Proceso de planeación de la demanda en el contexto del proyecto.....	11
Métodos de análisis de series de tiempo en la definición de pronósticos de demanda.....	12
Nuevo alcance definido para el proyecto.....	13
Capítulo 4 Fundamentación teórica.....	14
Estado del arte.....	14
Marco teórico.....	18
Capítulo 5 Metodología.....	22
Descripción de la metodología.....	22
Fases de la metodología.....	23
Fase I - Comprensión del negocio o problema.....	23
Fase II - Entendimiento de los datos.....	24
Fase III - Preparación de los datos.....	25
Fase IV - Modelado.....	25
Fase V - Evaluación.....	25
Fase VI - Despliegue.....	25
Resumen de tareas por fases en la metodología.....	25
Consideraciones.....	26
Capítulo 6 Solución propuesta.....	28
Descripción general de la solución.....	28
Diagrama de arquitectura de la solución.....	28
Estrategia para la definición de los pronósticos de ventas.....	30
Ventas General.....	30
Ventas por Asesor Comercial.....	30
Ventas por Cliente.....	31
Ventas por Producto.....	31
Información de software y hardware utilizado.....	32
Hardware.....	32
Software.....	32
Capítulo 7 Datos.....	34
Fuente de datos.....	34
Selección de las fuentes de datos.....	34
Diseño del modelo de datos.....	35

Limpieza y procesamiento de los datos.....	35xi
Transformación de los datos.....	36
Análisis exploratorio de datos .....	37
Estructura del set de datos .....	37
Resumen del set de datos.....	38
Tipo de datos para las variables del set de datos .....	39
Vista preliminar de datos.....	39
Visualización básica de datos .....	42
Preparación de datos .....	47
Tratamiento de valores perdidos (Missing Values) .....	47
Tratamiento de atípicos .....	50
Capítulo 8 Selección del modelo .....	62
Descripción general del proceso de modelado .....	62
Calibración de modelos.....	65
Cargar datos pre-procesados.....	65
Seleccionar grupos de modelos a calibrar .....	65
Modelos seleccionados para el análisis .....	66
Realizar muestreo de datos .....	68
Análisis de serie de tiempo para los datos seleccionados.....	69
Calibrar modelos por grupo.....	74
Medición del desempeño de los modelos o performance.....	82
Capítulo 9 Medición del desempeño de los Modelos .....	83
Esquema general.....	83
Carga de modelos calibrados.....	84
Realizar pronósticos con modelos calibrados.....	84
Medición del desempeño de modelos calibrados.....	84
Medir desempeño por grupo de modelos .....	84
Resumen de desempeño por grupo de modelos.....	84
Obtener el mejor modelo regresión lineal con y sin estacionalidad (LR).....	85
Obtener el mejor modelo autorregresivo (AR).....	86
Obtener el mejor modelo media móvil integrados autorregresivos estacionales (ARIMA) .....	87
Obtener el mejor modelo Holt-Winters (HW).....	88
Obtener el mejor modelo Feed-Forward Neural Networks and Multinomial Log-Linear Models (NNet) .....	89
Obtener el mejor modelo Recurrent Neural Network Elman and Jordan .....	91
Resumen de desempeño de modelos.....	95
Resumen del desempeño de grupo de modelos con tiempo de calibración .....	95
Desempeño para cada uno de los modelos.....	96
Top 5 de modelos con el mejor desempeño.....	99
Comparativo pronósticos modelos con mejor desempeño .....	99
Modelo con el mejor desempeño .....	100
Capítulo 10 Estrategia de implementación del modelo .....	102
Capítulo 11 Productos generados .....	106
Detalle de productos generados .....	107
Interface para la generación de pronósticos.....	107
Librerías para la generación automática de pronósticos en lenguaje R .....	107

Interface para la importación automática de resultados de pronósticos a la bodega de datos Corporativo (DWH) .....	xii
Interface para la consulta de pronósticos.....	108
Interface para la definición del plan de ventas con base en juicio de experto.....	109
Interface para el seguimiento al cumplimiento de presupuesto .....	113
Interface para la validación de asertividad de los pronósticos .....	115
Capítulo 12 Ventajas y desventajas en la adopción del modelo.....	116
Ventajas de la adopción del modelo .....	117
Desventajas de la adopción del modelo .....	117
Capítulo 13 Resultados obtenidos.....	118
Beneficios para la organización.....	118
Cumplimiento de los objetivos .....	119
Caracterización de la situación actual del proceso de planificación de la demanda.....	119
Automatización del proceso de generación de pronósticos de ventas .....	119
Ventajas y desventajas de la adopción del modelo de pronóstico de la demanda .....	121
Definición e implementación de la estrategia para la adopción del modelo .....	121
Tiempo de calibración y ejecución de Pronósticos.....	122
Ventas General.....	122
Ventas por asesor comercial .....	123
Ventas por Cliente.....	123
Ventas por producto .....	124
Capítulo 14 Conclusiones.....	126
Técnicas.....	126
Negocio .....	127
Capítulo 15 Trabajos futuros.....	129
Perspectiva de negocio.....	129
Integración de los pronósticos en los procesos de gestión .....	129
Perspectiva tecnológica.....	130
Administración de modelos predictivos .....	130
Lista de referencias .....	132
Apéndice.....	134

Tabla 1 Tipo de datos para las variables del set de datos.....	39
Tabla 2 Vista previa de los diez primeros registros del set de datos .....	40
Tabla 3 Top 10 del importe total de ventas por cliente.....	41
Tabla 4 Top 10 del importe total de ventas por asesor comercial .....	41
Tabla 5 Top 10 del importe total de ventas por producto .....	42
Tabla 6 Resumen de valores perdidos para las 25 variables del set de datos .....	49
Tabla 7 Consolidado de anomalías para la variable importe.....	59
Tabla 8 Consolidado de anomalías para la variable cantidad .....	61
Tabla 9 Tiempo de calibración de modelos regresión lineal con y sin estacionalidad .....	74
Tabla 10 Tiempo de calibración de modelos Autorregresivos (AR) .....	75
Tabla 11 Tiempo de calibración de modelos media móvil integrados .....	76
Tabla 12 Tiempo de calibración de modelos Holt-Winters (HW) .....	76
Tabla 13 Tiempo de calibración de modelos NNet .....	77
Tabla 14 Tiempo de Calibración de Modelos Recurrent Neural Network Elman and Jordan Neural.....	77
Tabla 15 Tiempo de calibración de modelos Extreme Gradient Boosting (XGBoost) .....	78
Tabla 16 Resumen de tiempo de calibración general de modelos.....	80
Tabla 17 Desempeño de modelos de regresión lineal con y sin estacionalidad (LR) .....	85
Tabla 18 Desempeño de modelos autorregresivos (AR).....	86
Tabla 19 Desempeño de modelos de media móvil integrados autorregresivos estacionales (ARIMA).....	88
Tabla 20 Desempeño de modelos Holt-Winters (HW).....	89
Tabla 21 Desempeño de modelos Feed-Forward Neural Networks and Multinomial Log-Linear Models.....	90
Tabla 22 Desempeño de modelos Recurrent Neural Network Elman and Jordan Neural Network .....	92
Tabla 23 Desempeño de modelos Extreme Gradient Boosting (XGBoost).....	95
Tabla 24 Desempeño general por grupo de modelos.....	96
Tabla 25 Desempeño general de modelos.....	97
Tabla 26 Top 5 de modelos con el mejor desempeño.....	99
Tabla 27. Estadísticas generales de ejecución de pronósticos de venta general .....	122
Tabla 28. Estadísticas generales de ejecución de pronósticos de ventas por asesor comercial ..	123
Tabla 30. Estadísticas generales de ejecución de pronósticos de ventas por producto.....	125

Figura 1. Contexto general del proyecto, de acuerdo con la estrategia para la adopción del modelo establecida. Elaboración propia. ....	10
Figura 2. Proceso general de planeación de la demanda actual en la organización. Elaboración propia.....	11
Figura 3. Proceso de planeación de la demanda en el contexto alcance del proyecto. Elaboración propia.....	12
Figura 4. Métodos de pronósticos de demanda definidos. Elaboración propia.....	12
Figura 5. Consolidación del alcance del proyecto por fases. Elaboración propia. ....	13
Figura 6. Fases de la metodología CRISP-DM en el contexto del proyecto y de la organización. Elaboración propia. ....	23
Figura 7. Fases de la metodología CRISP-DM. Elaboración propia. ....	23
Figura 8. Resumen de tareas por fases de la metodología CRISP-DM. Elaboración propia. ....	26
Figura 9. Diagrama de arquitectura de la solución. Elaboración propia.....	29
Figura 10. Diagrama de flujo general de la solución. Elaboración propia.....	30
Figura 11. Información de hardware utilizado. Elaboración propia. ....	32
Figura 12. Esquema general de las fuentes de datos. Elaboración propia. ....	34
Figura 13. Modelo de datos esquema en estrella del origen de datos. Elaboración propia. ....	35
Figura 14. Proceso ETL de limpieza y procesamiento de los datos de ventas. Elaboración propia. ....	36
Figura 15. Proceso ETL de transformación de datos de ventas. Elaboración propia. ....	37
Figura 16. Importe de ventas por fecha. Elaboración propia. ....	43
Figura 17. Importe de ventas por año. Elaboración propia. ....	43
Figura 18. Importe de ventas por mes. Elaboración propia. ....	44
Figura 19. Importe de ventas por día de la semana. Elaboración propia. ....	44
Figura 20. Importe de ventas por asesor comercial. Elaboración propia.....	45
Figura 21. Importe de ventas por clase de producto. Elaboración propia.....	45
Figura 22. Importe de ventas por gama de producto. Elaboración propia. ....	46
Figura 23. Descuento de línea de factura por año. Elaboración propia. ....	46
Figura 24. Descuento de línea de factura por mes. Elaboración propia.....	47
Figura 25. Histograma de valores perdidos. Elaboración propia. ....	48
Figura 26. Serie de tiempo importe de ventas general. Elaboración propia.....	50
Figura 27. Distribución general del importe de ventas con transformación logarítmica. Elaboración propia. ....	51
Figura 28. Serie de tiempo de venta de cantidad de productos. Elaboración propia.....	51
Figura 29. Distribución general de cantidades de productos vendidos con transformación logarítmica. Elaboración propia.....	52
Figura 30. Importe atípico consolidado por año. Elaboración propia.....	52
Figura 31. Importe atípico consolidado por mes. Elaboración propia. ....	53
Figura 32. Cantidades atípicas consolidadas por año. Elaboración propia. ....	53
Figura 33. Cantidades atípicas consolidadas por mes. Elaboración propia. ....	54
Figura 34. Test de Grubbs para la identificación de valores atípicos para importe y cantidad. Elaboración propia. ....	54
Figura 35. Histograma del importe de ventas. Elaboración propia. ....	55
Figura 36. Histograma del importe de ventas con tratamiento de atípicos. Elaboración propia. ...	55

Figura 37. Distribución del importe de ventas. Elaboración propia. ....	56xv
Figura 38. Distribución del importe de ventas con tratamiento de atípicos. Elaboración propia..	56
Figura 39. Histograma de cantidades. Elaboración propia.....	57
Figura 40. Histograma de cantidades con tratamiento de atípicos. Elaboración propia. ....	57
Figura 41. Distribución de cantidades. Elaboración propia. ....	57
Figura 42. Distribución de cantidades con tratamiento de atípicos. Elaboración propia.....	58
Figura 43. Detección de anomalías del importe de ventas por fecha. Elaboración propia. ....	59
Figura 44. Detección de anomalías para la variable cantidad. Elaboración propia. ....	60
Figura 45. Diagrama del proceso general de calibración de modelos. Elaboración propia. ....	63
Figura 46. Flujo del proceso general de calibración de modelos. Elaboración propia. ....	63
Figura 47. Comparativo del importe de ventas anual. Elaboración propia. ....	70
Figura 48. Serie de tiempo ventas general. Elaboración propia. ....	70
Figura 49. Serie de tiempo datos de entrenamiento ventas general. Elaboración propia. ....	71
Figura 50. Serie de tiempo datos de prueba ventas general. Elaboración propia. ....	71
Figura 51. Descomposición estacional de la serie de tiempo ventas general por Loess. Elaboración propia. ....	71
Figura 52. ACF Serie de tiempo datos de entrenamiento. Elaboración propia. ....	72
Figura 53. PACF Serie de tiempo datos de entrenamiento. Elaboración propia.....	73
Figura 54. ACF Serie de tiempo ventas general. Elaboración propia.....	73
Figura 55. PACF Serie de tiempo ventas general. Elaboración propia.....	73
Figura 56. Peridiograma de la serie de tiempo ventas general. Elaboración propia.....	74
Figura 57. Esquema general proceso de medición del desempeño. Elaboración propia. ....	83
Figura 58. Pronósticos regresión lineal con y sin estacionalidad (LR). Elaboración propia. ....	85
Figura 59. Pronósticos modelo autorregresivos (AR). Elaboración propia. ....	86
Figura 60. Pronóstico media móvil integrados autorregresivos estacionales (ARIMA). Elaboración propia. ....	87
Figura 61. Pronóstico Holt-Winters (HW). Elaboración propia.....	89
Figura 62. Pronóstico Modelos Feed-Forward Neural Networks and Multinomial Log-Linear Models (NNet). Elaboración propia.....	90
Figura 63. Pronóstico con datos de prueba modelo Elman, con parámetros básicos. Elaboración propia.....	91
Figura 64. Pronóstico con datos de prueba modelo Jordan, con parámetros básicos. Elaboración propia.....	92
Figura 65. Pronósticos con modelos Extreme Gradient Boosting (XGBoost). Elaboración propia. .....	93
Figura 66. Comparativo de pronósticos con modelos Extreme Gradient Boosting (XGBoost). Elaboración propia. ....	93
Figura 67. Comparativo de pronósticos de modelos con el mejor desempeño. Elaboración propia. .....	100
Figura 68. Comparativo de pronósticos de modelos con mejor el desempeño y consolidado. Elaboración propia. ....	100
Figura 69. Vista general de la estrategia de implementación. Elaboración propia.....	102
Figura 70. Interface de consulta de pronósticos de ventas. Elaboración propia.....	103
Figura 71. Graficas con detalle de histórico de ventas vs pronósticos ventas. Elaboración propia. .....	103
Figura 72. Interface de definición del presupuesto de ventas. Elaboración propia. ....	104

Figura 73. Proceso de planeación de la demanda propuesto. Elaboración propia.....	105xvi
Figura 74. Vista general de los productos generados por el proyecto. Elaboración propia.....	106
Figura 75. Productos generados por el proyecto, en orden de implementación. Elaboración propia.....	107
Figura 76. Interface para la generación de pronóstico. Elaboración propia.....	107
Figura 77. Estructura general de librerías del proyecto en R. Elaboración propia.....	108
Figura 78. ETL de carga de pronósticos a la bodega de datos corporativo. Elaboración propia.	108
Figura 79. Estructura del archivo de pronóstico generado. Elaboración propia.....	109
Figura 80. Vista de datos de pronósticos registrados en la bodega de datos. Elaboración propia. .....	109
Figura 81. Interface de consulta de pronósticos de ventas general. Elaboración propia. ....	110
Figura 82. Interface de consulta grafica de pronósticos de ventas general. Elaboración propia.	110
Figura 83. Interface de consulta de pronósticos por asesor comercial. Elaboración propia. ....	111
Figura 84. Interface de consulta grafica de pronósticos por asesor comercial. Elaboración propia. .....	111
Figura 85. Interface de consulta de pronósticos por cliente. Elaboración propia.....	112
Figura 86. Interface de consulta grafica de pronósticos por cliente. Elaboración propia. ....	112
Figura 87. Interface de consulta de pronósticos por producto. Elaboración propia. ....	113
Figura 88. Interface de consulta grafica de pronósticos por producto. Elaboración propia.....	113
Figura 89. Interface definición de presupuesto por juicio de experto. Elaboración propia. ....	114
Figura 90. Interface de selección de factores de impacto. Elaboración propia. ....	114
Figura 91. Interface de resumen de juicio de experto. Elaboración propia.....	115
Figura 92. Interface de seguimiento al cumplimiento del presupuesto. Elaboración propia. ....	115
Figura 93. Interface para la validación de la asertividad de los pronósticos. Elaboración propia. .....	116
Figura 94. Pronósticos de venta general. Elaboración propia. ....	119
Figura 95. Pronósticos de ventas por asesor comercial. Elaboración propia. ....	120
Figura 96. Pronósticos de ventas por cliente. Elaboración propia.....	120
Figura 97. Pronósticos de ventas de cantidades de productos. Elaboración propia. ....	121
Figura 98. Trabajos futuros desde la perspectiva de negocio. Elaboración propia. ....	130
Figura 99. Trabajos futuros desde la perspectiva del área de tecnología. Elaboración propia. ..	131

## **Capítulo 1**

### **Introducción y consideraciones generales**

#### **Introducción**

El presente trabajo de grado busca la definición de un modelo de pronóstico de series de tiempo basado en técnicas de analítica predictiva, en la mejora del proceso de definición del plan y presupuesto de ventas en la organización objeto de estudio. Se realiza el planteamiento de la solución con base en los objetivos y a las prioridades definidas por el negocio y la consolidación del alcance del proyecto en tres fases, se implementará únicamente la fase I del proyecto. La metodología que se adoptará será la CRISP-DM (Cross Industry Standard Process for Data Mining), es el estándar para la gestión de proyectos de análisis de datos ampliamente utilizada. CRISP-DM está conformada por un ciclo de seis etapas las cuales incluyen la comprensión del negocio, la comprensión de los datos, la preparación de los datos, el modelado y finalmente la evaluación del desempeño y la puesta en producción o despliegue. Para la selección del modelo, de acuerdo con la metodología, y con el fin de validar los modelos de pronósticos de la demanda basados en técnicas de analítica predictiva aplicados con éxito en la industria, se evalúan 36 modelos distribuidos en 7 grupos. La solución propuesta busca la automatización del proceso de generación de ventas para la definición del presupuesto y consolidación del plan de ventas como parte del proceso de planeación general de la demanda en la organización. Se analizarán las ventajas y desventajas de la adopción del y finalmente se implementará un sistema de información que permita a la organización la adopción del modelo y la automatización de los procesos establecidos en la estrategia.

## **Consideraciones generales**

La información contenida en el presente trabajo de grado es de carácter confidencial o reservada y tiene como fin único el ejercicio académico aquí propuesto. Queda expresamente prohibido copiar, distribuir o divulgar dicha información.

## **Capítulo 2**

### **Antecedentes**

#### **Planteamiento del problema**

##### **Enunciado del problema**

La internacionalización, la mundialización, la globalización, las nuevas tecnologías y en general la nueva economía son elementos que han creado un mercado cambiante, dinámico, complejo, lleno de incertidumbre, altamente competitivo y selectivo. Todo ello implica que las organizaciones y en especial las organizaciones empresariales estén obligadas a cambiar rápidamente con el fin de brindar respuestas oportunas al entorno.

Así mismo, los enormes cambios tecnológicos que se han venido presentando en la últimas décadas, tales como, el incremento exponencial de la capacidad de procesamiento computacional, la disminución de los costos de almacenamiento de datos, la aparición de tecnologías que facilitan el análisis de grandes volúmenes de datos, el perfeccionamiento de las técnicas analíticas, el aprendizaje de máquina, sistemas de gestión más adecuados y avanzados y en general la organización y la aplicación del conocimiento con fines prácticos, han venido transformando la forma en la cual las organizaciones toman decisiones basadas en datos e información (Kast y Rosenzweig, 1987). Dicha información se constituye en una forma eficaz de gestionar los recursos disponibles y de optimizar los procesos clave de la cadena de valor, con el fin de asegurar el cumplimiento de los objetivos propuestos y de mantener las ventajas competitivas (Cuatrecasas, 2012).

Ahora, la planificación de la demanda en la organización es el proceso que permite la toma de decisiones administrativas acerca de cómo hacer un uso eficiente de los recursos, los cuales permitan responder mejor a las previsiones de demandas futuras. Establecer pronósticos con mayor exactitud se constituye en un proceso crucial para el establecimiento de los planes de negocio, los planes de presupuestos, elaborar la visión más probable de lo que será la demanda futura, y de esta manera la toma de decisiones bajo un nivel de incertidumbre consensuado (Krajewski, 2008).

En este sentido, para la organización objeto de estudio, el pronóstico general de la demanda es un proceso que inicia en el área comercial, quienes establecen sus pronósticos de ventas (consolidado de ventas mensual) con base en los históricos de ventas con información de los dos últimos años, mediante el uso de métodos tradicionales (promedios, media, etc.), utilizando macros de Excel y con un alto grado de subjetividad con base en juicio de expertos en la toma de decisiones. Para el proceso de definición de pronósticos de ventas, no se contemplan patrones de demanda tales como tendencias, políticas de descuentos vigentes, ciclo de vida del producto, estacionalidad, procesos de licitación en curso, oportunidades de venta en curso y pérdidas de ventas por la no disponibilidad de inventario.

Además, dichos pronósticos se constituyen en insumo para los demás procesos de la cadena de suministro. El área de compras por ejemplo establece sus pronósticos con base en el histórico de compras con un catálogo de productos y maestro de materiales con un

número creciente de registros y contemplando la información de consumos de manera consolidada. En el proceso de planeación de los niveles de producción establecen sus pronósticos de la demanda de forma independiente (no se contemplan los pronósticos de compras, el ciclo de vida del producto, los niveles de inventario, la planeación de recursos, las necesidades de contratación de talento humano, de capacitación, etc.), basados en históricos de producción, con métodos tradicionales y con procesos ejecutados manualmente.

Al respecto, en un estudio inicial del proceso para la definición del pronóstico de ventas, se establece que el número de clientes es de 3346, de los cuales 3180 se encuentran en estado activo, 137 en estado inactivo y 29 son clientes potenciales, se ha presentado un incremento del 16.79% en clientes nuevos con relación al año anterior. El número de partners (proveedores) es de 2601, de los cuales 2583 se encuentran en estado activo y corresponden principalmente a compras en el exterior, compras nacionales y compra de suministros y 18 en estado inactivo. El catálogo de productos está compuesto por 10451 artículos (Las últimas décadas han visto a las compañías aumentar rápidamente su variedad de productos.).

Además, la organización cuenta con 23 almacenes propios y 66 en consigna. El número de facturas generadas en el 2015 fue de 24.195 y en el 2016 fue de 20.363 (Una factura está compuesta por una o más líneas de factura, cada línea obedece a una referencia de artículo, algunos artículos están formados por varios componentes). La variación en el

pronóstico general de ventas para el año 2016 fue del 27,23%. El pronóstico de ventas por Clientes y por asesor comercial, se establece con base en los históricos de los dos últimos años), mediante plantillas de Excel, utilizando modelos de regresión lineal simples, no se evalúa la efectividad de los resultados obtenidos con el modelo predictor (no se evidencian cifras). En general, este es un proceso que se realiza de forma manual (la definición del pronóstico toma en promedio dos meses), y el juicio de experto influye en gran medida el resultado final de la predicción y del pronóstico. Del análisis realizado para el cálculo de cantidad sugerida de artículos, se contempla la información de promedio mensual, trimestral y semestral del stock físico, stock bloqueado, stock reservado y stock bajo pedido, solo se tienen en cuenta los productos relacionados con el proceso de inyección, está planeado incluir en el análisis los productos para metalmecánica y tapicería. Al acceder a las plantillas para el cálculo de la cantidad sugerida, no se evidencian cifras de evolución de efectividad del modelo utilizado para la predicción. Finanzas define flujos de caja con base en la información definida por el área de compras y no se evidencian cifras de evaluación del modelo de pronóstico utilizado. No se contempla para el pronóstico de la demanda la información de mercadeo y finanzas. La planeación de la demanda se establece como el consenso de los gerentes de ventas, mercadeo, comercial, compras, producción y logística, con base en la información de pronósticos de la demanda anteriormente mencionados.

La desarticulación de los procesos clave de la cadena de valor para la planificación de la demanda y la toma de decisiones administrativas basada en pronósticos ejecutados de

la manera antes mencionada, repercuten en el incremento de la dificultad de administración general de los procesos, inventarios inexactos en almacenes y bodegas, niveles bajos de eficacia en la cadena de suministro, baja optimización de los inventarios, inventarios obsoletos, ineficiente control físico de inventarios, deterioro de insumos en el almacén y bodegas, pérdida de oportunidades de ventas al no mantener un adecuado stock de inventario (stockout), deficiente programación de compras, producción no ajustada a las necesidades reales de la organización, calidad deficiente en producto terminado, entrega inoportuna de información a clientes y proveedores, y en general la disminución de los índices de rentabilidad.

### **Formulación del problema**

En virtud de lo anterior, el presente trabajo de grado se encamina en la búsqueda de respuestas a los siguientes interrogantes:

- ¿Qué estrategias, metodologías, enfoques y modelos para la predicción de la demanda, a través de pronósticos con análisis de series de tiempo, se han adoptado con éxito en la industria?
- ¿Cuál modelo de pronóstico de la demanda basado en técnicas de analítica predictiva debería ser adoptado por la organización para la definición del plan y presupuesto de ventas?
- ¿Cuáles son las ventajas y desventajas de la adopción de un modelo de pronóstico de la demanda basado en técnicas de analítica predictiva para la definición del plan y presupuesto de ventas en la organización?

- ¿Qué estrategia debería ser adoptada por la organización para la integración del modelo de pronóstico de la demanda en la definición del plan y presupuesto de ventas?
- ¿Pueden las pequeñas y medianas empresas en Colombia con presupuestos ajustados, hacer uso de la analítica predictiva como estrategia de mejora del proceso de toma de decisiones y respuesta rápida al cambio?

## **Objetivos**

### **Objetivo general**

Definir el modelo más adecuado para el pronóstico de la demanda basado en técnicas de analítica predictiva, en la definición del plan y el presupuesto de ventas en la organización.

### **Objetivos específicos**

- Caracterizar la situación actual del proceso de planificación de la demanda e identificar las principales deficiencias que se presentan en el proceso de definición del plan y presupuesto de ventas, relacionados con los pronósticos de la demanda.
- Identificar los modelos de pronósticos de la demanda basados en técnicas de analítica predictiva aplicados con éxito en la industria.
- Analizar las ventajas y desventajas de la adopción del modelo de pronóstico de la demanda basado en técnicas de analítica predictiva para la organización.

- Definir una estrategia para la adopción del modelo para el pronóstico de la demanda en la organización.

## Capítulo 3

### Delimitación del alcance del proyecto

El alcance del proyecto inicialmente definido contemplaba la generación de pronósticos de la demanda para la gestión efectiva de compras e inventarios. Una vez presentando el proyecto a las directivas y líderes de procesos de la organización, y de definir las prioridades relacionadas con la planeación de la demanda, se establece que el nuevo alcance para el proyecto es la selección de un modelo de pronósticos para la definición del presupuesto y consolidación del plan de ventas. La información del presupuesto de ventas será un criterio de entrada a los procesos de planeación de demanda relacionados con la gestión de compras e inventarios, la cual será contemplada en una fase posterior al proyecto al interior de la Organización. La Figura 1, muestra el contexto general del proyecto con el nuevo alcance definido, las metodologías utilizadas para la gestión del proyecto y la estrategia para la adopción del modelo desde la perspectiva de procesos.



Figura 1. Contexto general del proyecto, de acuerdo con la estrategia para la adopción del modelo

establecida. Elaboración propia.

## Proceso general de planeación de la demanda

Para la definición de los objetivos comerciales, y con el fin de dar visibilidad a los procesos (áreas), el proceso general de planeación de la demanda en la organización inicia con la definición de las estrategias de negocio por parte de mercado y los planes comerciales. En el comité se define el plan de demanda, con base en los pronósticos y se ajustan de acuerdo con el juicio de experto, analogía histórica, investigación de mercado, etc. Y por ultimo se establece un consenso por parte del comité para la definición a alto nivel del plan de ventas (ver Figura 2).



Figura 2. Proceso general de planeación de la demanda actual en la organización. Elaboración propia.

## Proceso de planeación de la demanda en el contexto del proyecto

La Figura 3, muestra la relación de las fases del proceso actual de planeación de la demanda, con las fases establecidas en la metodología CRIPS-DM, y de acuerdo con el alcance definido para el proyecto.

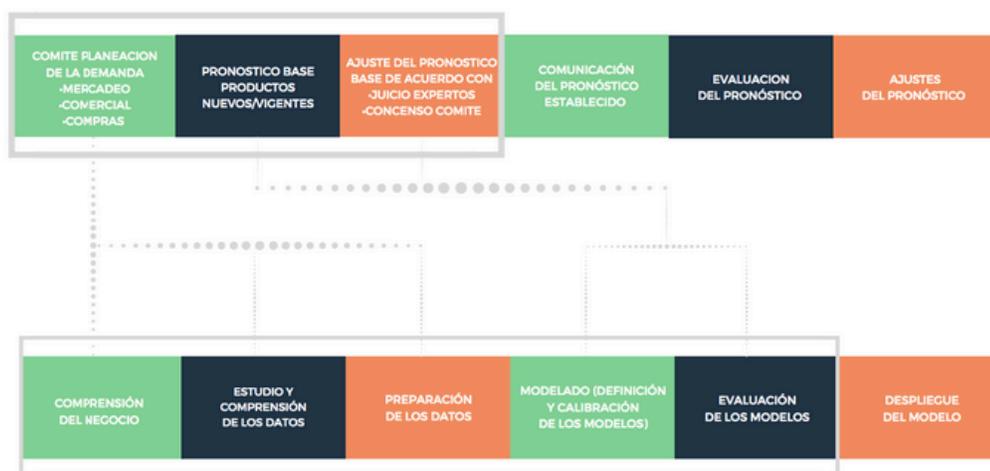


Figura 3. Proceso de planeación de la demanda en el contexto alcance del proyecto. Elaboración propia.

## Métodos de análisis de series de tiempo en la definición de pronósticos de demanda

Los enfoques definidos para el análisis de series de tiempo, en la definición del plan y presupuestó de ventas, incluyen métodos lineales, no lineales y métodos causales, mediante el uso de técnicas de analítica predictiva (regresión), específicamente técnicas de aprendizaje supervisado. Ver detalle en la Figura 4.



Figura 4. Métodos de pronósticos de demanda definidos. Elaboración propia.

## Nuevo alcance definido para el proyecto

El nuevo alcance definido para el proyecto es la selección de un modelo de pronóstico de la demanda basado en técnicas de analítica predictiva, en la definición del plan y el presupuesto de ventas en la organización. Los cuales están contemplados en la Fase I del proyecto. La integración de los pronósticos de la demanda, la planificación de la demanda, y la planificación integrada de la cadena de suministra serán contempladas en una dase posterior del proyecto y será gestionada por la organización (Ver Figura 5).



Figura 5. Consolidación del alcance del proyecto por fases. Elaboración propia.

## Capítulo 4

### Fundamentación teórica

#### Estado del arte

De acuerdo con Arango et al. (2013), para la estimación de la demanda futura de un producto se hace uso de técnicas de pronóstico (modelos) en el análisis de la demanda basados en información histórica cuantitativa y cualitativa. De esta manera los pronósticos de ventas pasan a ser una importante fuente de información para prever la demanda con un margen de incertidumbre concertado. Además, consideran que los pronósticos de ventas y la planeación de inventarios cobran importancia dado su impacto en los resultados económicos para la organización. Ahora, consideran que las investigaciones más recientes en esta área integran simultáneamente los pronósticos de ventas y la planeación de los inventarios, proponiendo, muchas veces, mejoras o alternativas a los algoritmos clásicos.

Murray et al. (2015), establecieron un modelo para el pronóstico de la demanda de la cadena de suministro, en entornos en donde existe un gran número de clientes y el cálculo de pronóstico no es manejable de forma individual. Mediante la utilización de técnicas de minería de datos identificaron segmentos de clientes (clústeres de clientes) con comportamientos de demanda similares. Una vez identificados los segmentos de clientes, permitieron la simplificación de la ejecución de los modelos de pronóstico para cada uno de los segmentos establecidos.

El modelo propuesto por Murray et al. (2015), podría ser implementado para el pronóstico de cantidades de productos de compra (en especial los de importación) con características y demandas similares. Además, podrían ser evaluados diferentes métodos de regresión y de análisis de series de tiempo, considerando la medición de error del pronóstico y evaluando el desempeño de los mismos.

De acuerdo con (Gardner, 2006; Taylor, 2006) citados por Du et al. (2016) el análisis de series de tiempo (TSF, por sus siglas en inglés) tradicional se ha realizado a través de métodos estadísticos, como los modelos de suavizado exponencial y los modelos de promedio móvil autorregresivo integrado-ARIMA (Contreras, Espínola, Nogales y Conejo, 2003) como modelos lineales. Mientras que, en las últimas décadas, el uso de las redes neuronales artificiales (RNA, por sus siglas en inglés) desempeña un papel determinante en la solución de problemas de TSF debido a un mayor desempeño en los problemas de regresión y de clasificación. Du et al. (2016) propusieron dos algoritmos evolutivos multi objetivos basados en RNA, para la solución de problemas de TSF. La principal diferencia que distingue a las RNA de los métodos tradicionales es la capacidad de generar relaciones no lineales ocultas en los datos de series de tiempo.

Wang y Xu (2014), propusieron un modelo Bayesiano de predicción el cual permite combinar métodos de predicción individuales de diferentes partes de la cadena de suministro para luego ser integrada a través de un modelo de planificación, pronóstico y reabastecimiento en colaboración (CPFR, por sus siglas en inglés). De acuerdo con

Guazzelli (2012), el proceso CPRF permite la administración de la cadena de valor en nueve pasos en el cual los pronósticos desempeñan una función determinante, pues permiten a una organización y a sus proveedores trabajar conjuntamente en la elaboración del pronóstico, mediante el uso compartido de la información tal como la demanda regular e irregular a través de internet. Consideramos que para la organización objeto de estudio, esta última parte del modelo propuesto por Wang y Xu (2014) no es viable, dado que aún no existe una cultura organización la cual permita compartir información de la demanda con los proveedores clave en la cadena de abastecimiento.

Aburto y Weber (2007), plantearon un sistema inteligente híbrido que combina los modelos de promedio móvil autorregresivo integrado (ARIMA) y las redes neuronales para el pronóstico de la demanda, con el fin de ayudar a mejorar la gestión de la cadena de suministro en la industria minorista. Analizando los resultados del estudio y con respecto a la precisión de los pronósticos, determinaron que las redes neuronales superan a los modelos ARIMA y el enfoque híbrido aditivo propuesto presentaron mejores resultados (mayor precisión en los pronósticos).

Carbonneau et al. (2008), establecieron un modelo para el pronóstico de la demanda al final de la cadena de suministro, mediante la utilización de técnicas avanzadas de aprendizaje mecánico (machine learning), tales como redes neuronales, redes neuronales recurrentes y máquinas de soporte vectorial. Compararon los resultados obtenidos con las técnicas de aprendizaje mecánico con enfoques tradicionales tales como Naive, análisis de

tendencia, media móvil, y regresión lineal, sugieren que, si bien las redes neuronales recurrentes y las máquinas vectoriales de soporte muestran el mejor rendimiento, su precisión de predicción no fue estadísticamente mejor que las del modelo de regresión. Consideramos que el rendimiento es fundamental en el procesamiento de pronósticos de demanda con un alto número de clientes y de productos, pero básicamente lo que se pretende es lograr una mayor precisión en los pronósticos de la demanda, por tanto, la utilización de estas técnicas de aprendizaje automático no será de prioridad en la selección, calibración y medición de la precisión del modelo.

Chen y Guestrin (2016), propusieron un método de aprendizaje automático a gran escala, altamente eficaz y ampliamente utilizado en problemas de clasificación y regresión mediante el uso de árboles de decisión escalables (Tree Boosting). Aprovechando el uso de la computación paralela XGBoost escala más allá de miles de millones de registros los cuales utilizan menos recursos que los métodos existentes. Este enfoque podría ser adoptado ante la necesidad de generación de pronósticos para cientos de productos dada la disponibilidad de recursos de cómputo de la organización.

Según Arango et al. (2013), los avances en las tecnologías de la información y de las comunicaciones, la dinámica de los mercados, la reducción de los tiempos de suministro, la globalización, han venido transformando la forma en la cual los modelos de inventarios puedan ser todos de revisión continua y agrega:

...Los modelos tradicionales como el del Lote Económico de pedido que calcula la cantidad óptima a pedir a partir de minimizar el costo total (compuesto básicamente por el costo de hacer un pedido y el costo de almacenamiento); han empezado a ser revaluados, teniendo en cuenta que pedir es cada vez menos costoso mientras que almacenar tiende a ser cada vez más riesgoso y oneroso para las organizaciones. (p. 743).

### **Marco teórico**

Es indudable que actualmente se está viviendo una época en que las empresas han cambiado la forma en la cual compiten, debido a factores tales como, mayor presión competitiva, creciente nivel de exigencia por parte de los clientes, el uso de tecnologías fácilmente reproducibles, la oferta de productos y/o servicios similares. Muchas de las fuentes previas para la generación de ventajas competitivas sustentables hoy en día ya no son aprovechables. Los procesos empresariales de alto rendimiento están entre los últimos aspectos de diferenciación que aún quedan. Con el fin de incrementar los niveles de eficiencia, eficacia y efectividad que permitan un mejor desempeño organizacional, es indispensable para la organización tomar las decisiones más inteligentes que le sea posible (Davenport & Harris, 2007). De acuerdo con Guazzelli (2012), la aparición de nuevas disciplinas como la inteligencia empresarial, nos ha permitido solucionar en gran medida la necesidad de entender eventos pasados, y de esta manera tomar decisiones con base en estadísticas obtenidas a partir de datos históricos. Sin embargo, estas técnicas simplemente no son suficientes. En el contexto de la sociedad de la información ha surgido la necesidad de tener mejores, más rápidos y eficientes métodos para extraer y transformar los datos de una organización en información y distribuirla a lo largo de la cadena de valor. Mediante

el uso de metodologías, aplicaciones, prácticas y capacidades enfocadas a la creación y administración de información que permite tomar mejores decisiones a los usuarios de una organización (Curto y Conesa ,2012).

Además, de acuerdo con Etxeberria (2005), dentro de la estadística existen dos grandes campos bien diferenciados los cuales son: la estadística descriptiva y la estadística inferencial. La estadística descriptiva es la parte de la disciplina que tiene como objetivo el recoger, organizar, resumir, describir y analizar un conjunto de datos mediante una serie de técnicas y métodos y de presentar los datos correspondientes, sin pretender ir más allá del propio conjunto de datos, en general nos permite saber qué sucedió en el pasado. La estadística inferencial por su parte tiene como objetivo el generalizar los resultados obtenidos en una muestra a la población objeto de estudio. Para poder llevar a cabo dicha generalización, se hace imprescindible el hacer uso de la teoría de la probabilidad.

Para Chapman (2006, p. 17), “La formulación de pronósticos (o proyección) es una técnica para utilizar experiencias pasadas con la finalidad de predecir expectativas del futuro.”. De esta forma, el pronóstico no es realmente una predicción, sino una proyección (con el horizonte de tiempo definido) estructurada de los datos, la información y el conocimiento pasado. Así mismo, para el desarrollo de modelos de pronóstico de la demanda se usan dos tipos generales de técnicas: los métodos cualitativos y los métodos cuantitativos. Entre los métodos cuantitativos se pueden encontrar: los métodos de juicio, encuestas de mercado, Delphi o consenso de panel, analogías de ciclo de vida y valoración

informada. Entre los métodos cuantitativos se pueden encontrar los métodos causales y el análisis de series de tiempo. Algunos de los enfoques más comunes de métodos causales son: los modelos de entrada-salida, modelos econométricos, modelos de simulación y de regresión. Finalmente, el análisis de series de tiempo como método cuantitativo se basa en el análisis de patrones de la demanda pasada de componentes tales como tendencia, estacionalidad, ciclicidad y la aleatoriedad y si estos patrones pueden ser analizados podrían ser utilizados para desarrollar proyecciones de la demanda futura, suponiendo que los patrones continúan aproximadamente de la misma forma (Chapman, 2006).

Para el desarrollo de modelos de pronóstico de la demanda, las compañías utilizan analítica predictiva, la cual tiene como objetivo general obtener un valor futuro bajo un umbral de incertidumbre concertado (tratando de encontrar un patrón de comportamiento) y de esta manera deducir cuál será (tomado como base a información histórica) el comportamiento en el futuro. Como disciplina, la analítica predictiva ha existido durante muchas décadas. Su relevancia en la industria se incrementó ante el crecimiento constante de la cantidad de datos producidos por los Clientes (por ejemplo, de transacciones online y redes sociales), el Internet de las cosas (por ejemplo, sensores, dispositivos GPS móviles, tecnologías RFID) y de las aplicaciones corporativas (por ejemplo, ERP, CRM, sistemas gerenciales), una mayor capacidad de almacenamiento a costos cada vez más asequibles ya sea basado en nubes públicas o privadas, y a una mayor disponibilidad de poder de procesamiento a través del uso de soluciones en su mayoría de código abierto y de uso libre (Guazzelli, 2012).

En síntesis, para el pronóstico de la demanda se hace uso de métodos cualitativos y cuantitativos. La analítica predictiva es la aplicación de técnicas matemáticas avanzadas tales como la estadística, las ciencias de la computación, la minería de datos o exploración de datos y el aprendizaje automático supervisado tales como máquinas de soporte vectorial (SVM, por sus siglas en inglés), redes neurales de retro propagación (backpropagation), K-vecinos más cercanos (K nearest neighbors) y árboles de decisión y no supervisado (entre los más usados se encuentra la agrupación en clúster), con el fin de intentar descubrir patrones y aprender de la correlación de los atributos en un conjunto de datos de entrada (datos actuales e históricos reales) y los datos de salida deseados. En otras palabras, intentar completar los datos de acontecimientos inciertos en un horizonte de tiempo determinado en el futuro.

El presente trabajo de grado tiene como fin definir las técnicas, evaluar los modelos utilizados con éxito en la industria para el pronóstico de la demanda (anteriormente mencionados) y la selección del modelo que mejor se ajusta a los datos históricos de compras e inventarios en la organización, de acuerdo con la metodología propuesta.

## Capítulo 5

### Metodología

#### Descripción de la metodología

El tipo de investigación que se llevará a cabo en el presente trabajo de grado será de estudio descriptivo, se hará uso de analítica predictiva para el pronóstico de la demanda, en la definición del plan y el presupuesto de ventas en la organización. Se contemplará información histórica de las operaciones relacionadas de los cinco últimos años, los horizontes de tiempo para los pronósticos de la demanda serán a corto (1 y 2 meses) y mediano (1 año) plazo. La información se obtendrá de los sistemas de información de la Compañía, principalmente del ERP (Órdenes de Venta, Facturación, Artículos, Gestión de Inventarios, Compra de artículos y de materias primas), el CRM (Información relacionada con el comportamiento de compra del Cliente), y el Data Warehouse Corporativo (Consolidado ventas). El número de registros que inicialmente se utilizarán en el trabajo de investigación serán aproximadamente 162.183 registros (filas) con 25 observaciones (columnas).

La metodología que se adoptará en el presente trabajo de investigación será CRISP-DM (Cross Industry Standard Process for Data Mining), es la guía de referencia ampliamente utilizada en el desarrollo de proyectos de minería de datos. Desarrollada por el consorcio compuesto inicialmente por Daimler Chrysler, SPSS. CRISP-DM es una metodología para la implementación de proyectos de análisis de datos (Minería de Datos), la cual está conformada por un ciclo de seis etapas las cuales son: Comprensión del

negocio, Comprensión de los datos, Preparación de los datos, Modelado, Evaluación y Despliegue (Santos & Azevedo, 2005).

La Figura 6, muestra las fases de la metodología en el contexto del proyecto y de la organización una vez definido el alcance de la fase I.

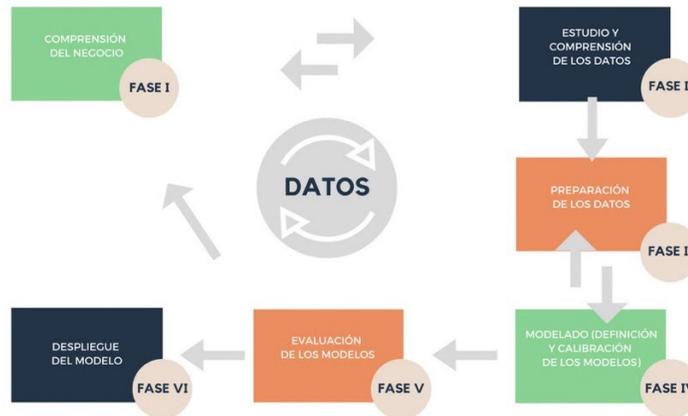


Figura 6. Fases de la metodología CRISP-DM en el contexto del proyecto y de la organización.

Elaboración propia.

## Fases de la metodología

La Figura 7, nos muestra de acuerdo con Putler & Krider (2012) las fases de la metodología para el desarrollo de proyectos en minería de datos CRISP-DM.



Figura 7. Fases de la metodología CRISP-DM. Elaboración propia.

### Fase I - Comprensión del negocio o problema

Se enfoca en la comprensión de los objetivos y de los requerimientos del proyecto desde la perspectiva de la organización.

Luego de la revisión del alcance inicial del proyecto (trabajo de grado), se determina que la prioridad para la organización en el proceso de planeación de la demanda es la definición del presupuesto y el plan de ventas.

### Determinar los objetivos del negocio

De acuerdo con el alcance definido para el proyecto (trabajo de grado), las áreas involucradas en el proceso de planeación de la demanda y específicamente en el proceso de definición del presupuesto y el plan de ventas son el área Comercial y Mercadeo. Una vez definido el alcance para el proyecto y de acuerdo con las prioridades, se establecen los siguientes objetivos:

- Automatizar el proceso de generación de pronósticos de ventas a través de métodos cualitativos
- Integrar los pronósticos de ventas a través de métodos cualitativos, al proceso de definición del Plan de Ventas
- Mejorar el proceso de toma de decisiones en la organización, con base en analítica predictiva
- Crear una cultura en torno a la toma de decisiones basados en los datos.

### **Fase II - Entendimiento de los datos**

El entendimiento o comprensión de datos implica la recolección inicial de datos, acceder a los datos, realizar una descripción de los datos, verificar la calidad de los datos y explorarlos con el fin de establecer mediante tablas, gráficos la forma en la cual se relacionan.

La recopilar los datos iniciales, se detalla en la sección Fuente de Datos. La descripción de los datos se detalla en la sección Análisis Exploratorio de Datos. La revisión de los datos se detalla en la sección preparación de datos y la verificación de la calidad de los datos, se detalla en la sección análisis exploratorio de datos

**Fase III - Preparación de los datos**

La preparación de datos incluye las tareas de integración de orígenes de datos, selección de datos, limpieza de datos, formateo de datos, agregación de datos, generación de nuevas variables y la partición de los datos en conjuntos de datos de entrenamiento y pruebas.

**Fase IV - Modelado**

En esta fase se incorporan los datos preparados para la construcción de los modelos de analítica predictiva y se evalúan con el fin de probar la efectividad de estos en la solución del problema planteado. Las tareas genéricas relacionadas en esta fase son: la seleccionar la técnica de modelado, generar el plan de prueba, construir el modelo y evaluar el modelo.

**Fase V - Evaluación**

En esta fase se evalúa el modelo, se selecciona el modelo (o modelos) que poseen un nivel de precisión adecuado para el cumplimiento de los criterios de éxito del problema establecido.

**Fase VI - Despliegue**

Una vez que el modelo ha sido construido y validado, se define la estrategia para incorporar el modelo desarrollado en los procesos empresariales relevantes para la organización.

**Resumen de tareas por fases en la metodología**

La Figura 8, nos muestra el resumen de las tareas a realizar para cada una de las fases propuestas en la metodología.

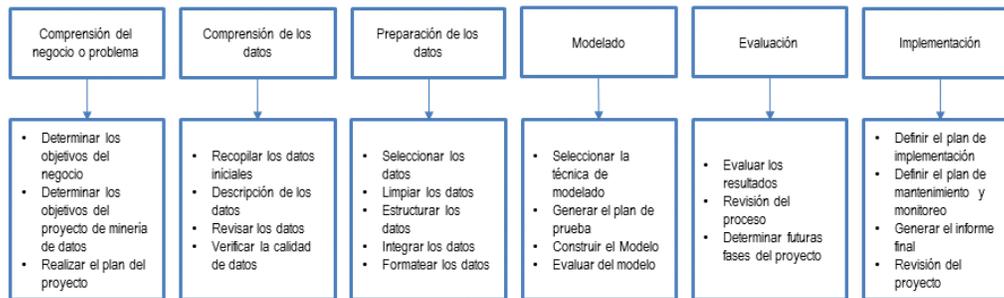


Figura 8. Resumen de tareas por fases de la metodología CRISP-DM. Elaboración propia.

## Consideraciones

Las fases cuatro y cinco de la metodología se han ejecutado recursivamente con el fin de seleccionar, calibrar y medir la precisión de los modelos y/o combinación de estos. Para la fase seis o de despliegue (puesta en producción), se realiza una propuesta para la integración del modelo seleccionado en la definición del plan y presupuesto de ventas, el cual forma parte del proceso general de planeación de la demanda.

De acuerdo con Vermorel (2013), en estadística, la precisión del pronóstico es el grado de cercanía (diferencia) entre el valor que se mide (expresión de cantidad) y el valor real de dicha cantidad. En otras palabras, la comparación entre el valor pronosticado y el valor real. El error del pronóstico podría establecerse como la diferencia entre la demanda real y el valor pronosticado. El cálculo del error del pronóstico de la demanda nos permitirá establecer que modelos presentan mayor precisión.

Las métricas utilizadas para la evaluación del desempeño (precisión) de los modelos para el pronóstico de venta general, ventas por asesor comercial, ventas por cliente y ventas por producto son:

- Mean Absolute Error (MAE)
- Mean Squared Error (MSE)
- Mean Absolute Percentage Error (MAPE).

Las métricas utilizadas en la medición y evaluación del desempeño de los modelos se detallarán en la sección de medición del desempeño.

## Capítulo 6

### Solución propuesta

#### Descripción general de la solución

Una vez establecidos los objetivos del negocio en la fase de comprensión del problema, y de acuerdo con las condiciones, alcance, prioridades y restricciones definidas para el proyecto, la solución general propuesta abarca las fases y mejores prácticas establecidas en la metodología CRISP-DM para la ejecución de proyectos de minería de datos y la definición de estrategia para la adopción del modelo a los procesos de negocio relacionados con la definición del plan y presupuesto de ventas.

#### Diagrama de arquitectura de la solución

Los componentes que forman parte de la arquitectura de una solución son las bases de datos de los sistemas ERP y CRM, los flujos de extracción, transformación y carga de datos a la bodega de datos corporativa y específicamente al datamart de ventas. La solución ETL empleada es Pentaho Data Integration. Las librerías y funciones desarrolladas en lenguaje R, permiten la automatización de los procesos de preparación de datos, el análisis de series de tiempo, el modelado, la medición del performance, la ejecución de pronósticos de acuerdo con el horizonte de tiempo definido y la visualización de resultados. Se ha desarrollado una interface web, la cual permite definir mediante juicio de expertos el presupuesto de ventas, los factores de impacto internos y externos (de manera general o específica) y las interfaces para la consolidación del plan de ventas, el seguimiento al cumplimiento del presupuesto y la medición de la asertividad de los pronósticos de ventas,

de acuerdo con las restricciones de acceso a la información, perfiles de usuario, niveles de aprobación establecidos por la organización (ver Figura 9).

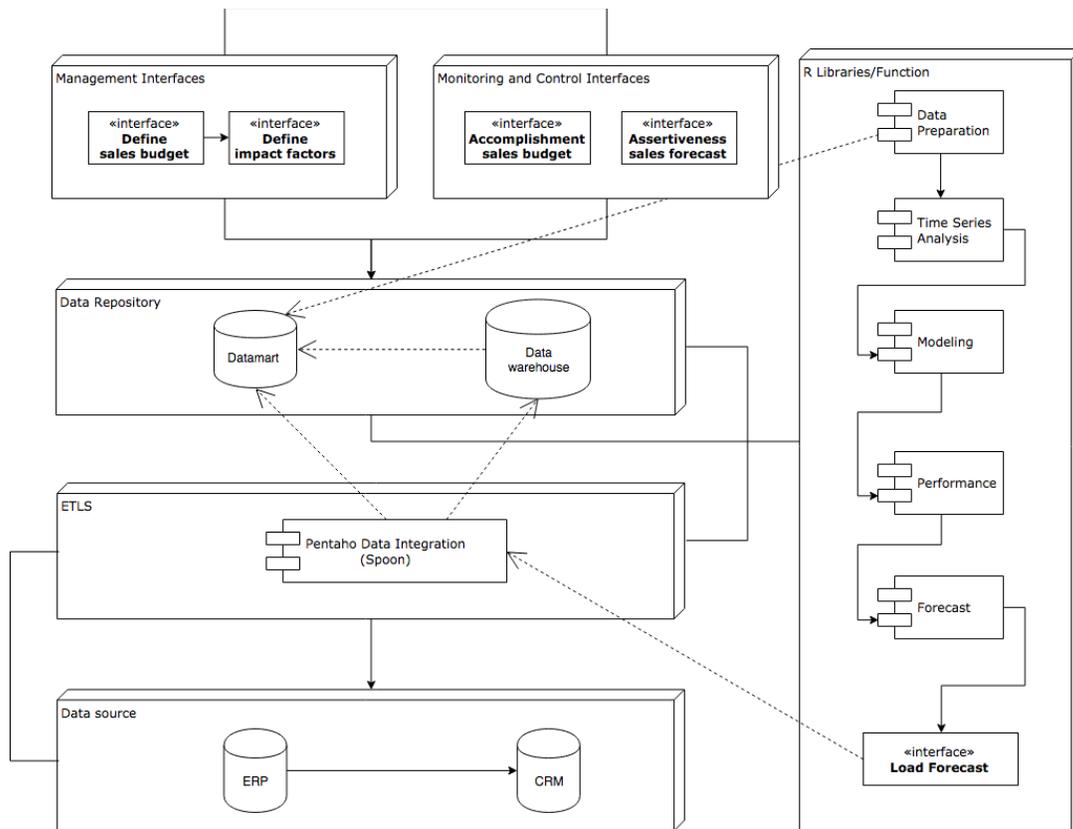


Figura 9. Diagrama de arquitectura de la solución. Elaboración propia.

La Figura 10, nos muestra el detalle del flujo de trabajo para la selección del modelo y la definición de la estrategia de adopción del modelo e integración a los procesos de negocio definidos.

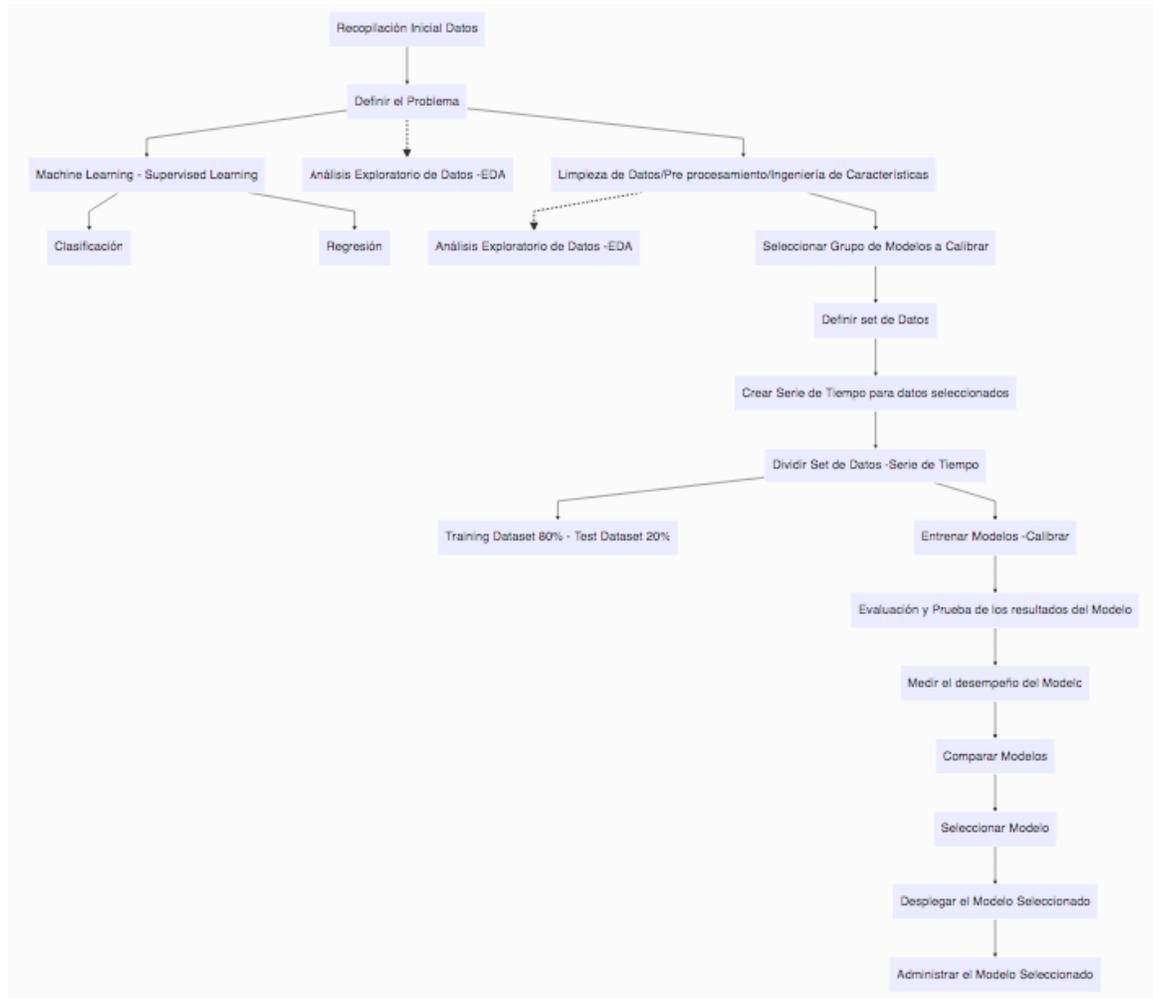


Figura 10. Diagrama de flujo general de la solución. Elaboración propia.

## Estrategia para la definición de los pronósticos de ventas

### Ventas General

Se realizan pronósticos de Ventas General para los siguientes 24 Periodos, con la información de histórico de ventas definido.

### Ventas por Asesor Comercial

Se realizan Pronósticos de Ventas para los 12 Asesores Comerciales y los siguientes 24 Periodos.

**Ventas por Cliente**

De los **3346** Clientes y de los **3180** activos, y con el fin de priorizar la ejecución de la generación de pronósticos para los siguientes 24 periodos, se aplica ley de Pareto en el cual se selecciona el 20% de los Clientes (**103 Clientes**) los cuales representan el 80% de las Ventas. Para los 3077 restantes, se pronosticarán con el modelo con el mejor desempeño y menor tiempo de ejecución.

**Ventas por Producto**

De los 10451 Productos y con el objetivo de priorizar la ejecución de la generación de pronósticos para los siguientes 24 periodos, se realiza la segmentación de artículos mediante el uso de la metodología de Clasificación ABC de inventarios (los criterios utilizados fueron: importe de ventas general y rentabilidad), así:

- **Categoría A** (Los más importantes): Total productos 522, los cuales corresponden al 5% del Total
- **Categoría B** (Con importancia media): Total productos 3135, los cuales corresponden al 30% del Total
- **Categoría C** (Poco importantes): Total productos 6794, los cuales corresponden al 65% del Total

Se priorizará la ejecución de pronósticos para los 522 productos que pertenecen a la Categoría A. Para los productos de las categorías B y C se pronosticarán con el modelo con el mejor desempeño y menor tiempo de ejecución.

## Información de software y hardware utilizado

Se ha utilizado un sistema Operativo Apple macOS High Sierra, un entorno de desarrollo para la definición de modelo R Studio, con las librerías y drivers para el acceso al repositorio de datos Corporativo.

### Hardware

Para la implementación de la solución propuesta, se ha empleado un equipo de computo MacBook Pro con las características de memoria y procesador detalladas en la Figura 11.



Figura 11. Información de hardware utilizado. Elaboración propia.

### Software

Resumen del Sistema Operativo, versión y librerías del lenguaje R utilizados, en la definición del modelo:

**R version 3.4.2 (2017-09-28)**

**Platform:** x86\_64-apple-darwin15.6.0 (64-bit)

**Locale:** en\_US.UTF-8|en\_US.UTF-8|en\_US.UTF-8|C|en\_US.UTF-8|en\_US.UTF-8

**Attached base packages:** stats, graphics, grDevices, utils, datasets, methods and base

**Other attached packages:** *DiagrammeR(v.0.9.2)* and *pander(v.0.6.1)*

**Loaded via a namespace (and not attached):** *Rcpp(v.0.12.16)*, *plyr(v.1.8.4)*, *compiler(v.3.4.2)*, *RColorBrewer(v.1.1-2)*, *influenceR(v.0.1.0)*, *highr(v.0.6)*, *bindr(v.0.1)*, *viridis(v.0.4.0)*, *tools(v.3.4.2)*, *digest(v.0.6.12)*, *jsonlite(v.1.5)*, *viridisLite(v.0.2.0)*, *gtable(v.0.2.0)*, *evaluate(v.0.10.1)*, *tibble(v.1.3.4)*, *rgexf(v.0.15.3)*, *pkgconfig(v.2.0.1)*, *rlang(v.0.1.2)*, *igraph(v.1.1.2)*, *rstudioapi(v.0.7)*, *yaml(v.2.1.16)*, *xfun(v.0.1)*, *bindrcpp(v.0.2)*, *gridExtra(v.2.3)*, *downloader(v.0.4)*, *dplyr(v.0.7.4)*, *stringr(v.1.3.0)*, *knitr(v.1.18)*, *htmlwidgets(v.0.9)*, *hms(v.0.3)*, *grid(v.3.4.2)*, *rprojroot(v.1.2)*, *glue(v.1.2.0)*, *R6(v.2.2.2)*, *Rook(v.1.1-1)*, *XML(v.3.98-1.9)*, *rmarkdown(v.1.8)*, *bookdown(v.0.6)*, *ggplot2(v.2.2.1)*, *tidyr(v.0.7.2)*, *purrr(v.0.2.4)*, *readr(v.1.1.1)*, *magrittr(v.1.5)*, *backports(v.1.1.1)*, *scales(v.0.5.0)*, *htmltools(v.0.3.6)*, *assertthat(v.0.2.0)*, *colorspace(v.1.3-2)*, *brew(v.1.0-6)*, *stringi(v.1.1.7)*, *visNetwork(v.2.0.1)*, *lazyeval(v.0.2.0)* and *munsell(v.0.4.3)*

## Capítulo 7

### Datos

#### Fuente de datos

El origen de datos para la definición del modelo de pronósticos de ventas son el ERP (Enterprise resource planning por sus siglas en inglés) o planificador de recursos empresariales y el CRM (Customer relationship management por sus siglas en inglés) o administración de las relaciones con el Cliente. Se realiza transformación de datos de ventas (normalizados) a través de ETLs para ser cargados a la bodega de datos Corporativa (DWH). De la bodega de datos se toma la información para la calibración y selección de los modelos con mejor desempeño. En esta fase se contemplan las tareas de integración de orígenes de datos y la selección de datos.

#### Selección de las fuentes de datos

Las fuentes de datos para el análisis de series de tiempo son los sistemas transaccionales ERP y CRM (ver Figura 12).

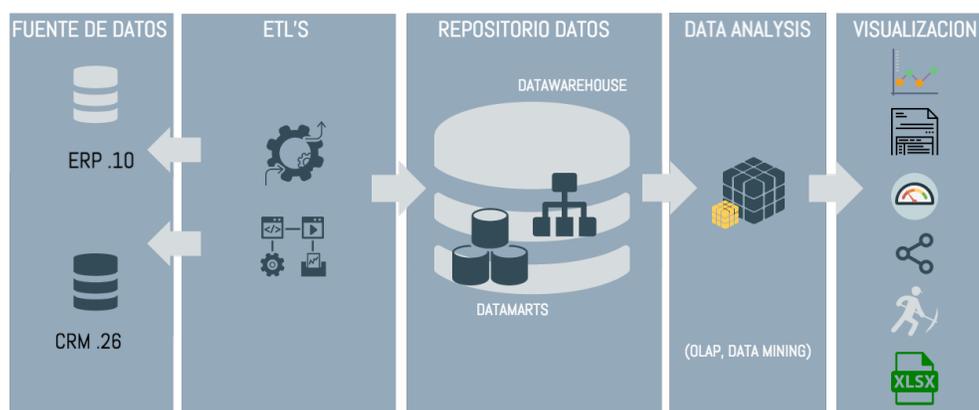


Figura 12. Esquema general de las fuentes de datos. Elaboración propia.

## Diseño del modelo de datos

La Figura 13, nos presenta una vista multidimensional de las tablas de hechos y dimensiones del Data Warehouse (Data Mart) relacionadas con el proceso de ventas y métricas del área comercial, las cuales serán utilizadas en la definición del presupuesto de ventas por asesor comercial, consolidación del plan de ventas y el seguimiento al cumplimiento del presupuesto.

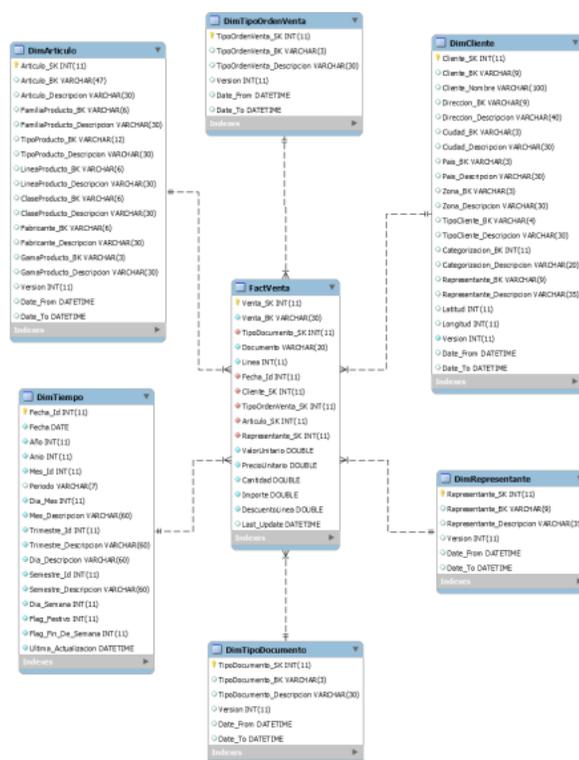


Figura 13. Modelo de datos esquema en estrella del origen de datos. Elaboración propia.

## Limpieza y procesamiento de los datos

Nos permite integrar los datos desde las fuentes relacionadas (orígenes de datos), cumpliendo con los estándares de calidad definidos y con la estructura establecida en el maestro de datos. Se realiza validación de facturas y documentos de ajuste, de acuerdo con las cifras reportadas a las entidades de control asegurando la no existencia de duplicidades y no diferencias en los totales de ventas mensuales y anuales. La Figura 14, nos muestra la

programación del Job en la herramienta Pentaho Data Integration para la importación de información de la tabla de hechos de ventas y dimensiones relacionadas.

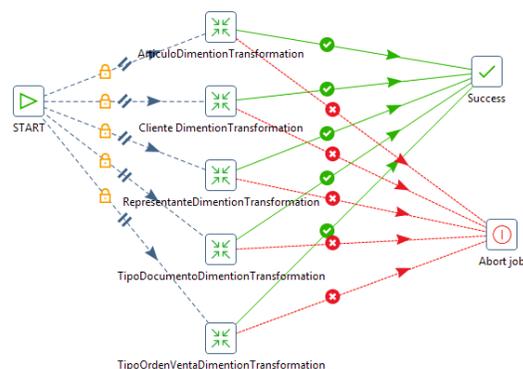


Figura 14. Proceso ETL de limpieza y procesamiento de los datos de ventas. Elaboración propia.

### Transformación de los datos

Nos permite aplicar una serie de reglas de negocio o funciones sobre los datos de origen para convertirlos en datos que serán finalmente cargados a la Bodega de Datos Corporativa (ver Figura 15). Para los asesores comerciales, categorización de Cliente, productos, clase de producto, gama de producto, línea de producto y tipo de producto no clasificados, se asigna la categoría NA (No Aplica), con el fin de asegurar la integridad de datos para el análisis para la definición de los pronósticos de ventas. Estos atributos podrían ser utilizados en la segmentación de Clientes y productos utilizando métodos de clusterización inicialmente planteados.

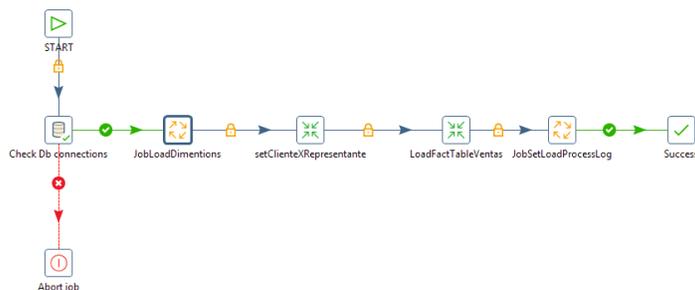


Figura 15. Proceso ETL de transformación de datos de ventas. Elaboración propia.

## Análisis exploratorio de datos

El Análisis Exploratorio de Datos o (E.D.A. por sus siglas en inglés) reúne un conjunto de técnicas estadísticas las cuales permiten explorar y comprender mediante tablas y gráficos la naturaleza y la forma en la cual los datos están relacionados.

### Estructura del set de datos

Las dimensiones del set de datos son:

- Numero de filas o registros: 162183.
- Numero de columnas o variables: 25.

Detalle del set de datos:

```
'data.frame': 162183 obs. of 25 variables:
 $ TipoDocumento : chr "V20" "V20" "V20" "V20" ...
 $ Documento : chr "10000913" "10000913" "10000913" "10000914" ...
 $ Linea : int 1 2 3 1 2 3 4 1 2 3 ...
 $ FechaId : int 20120101 20120101 20120101 20120101 20120101 20120101 20120101
20120101 20120101 20120101 ...
 $ Fecha : chr "2012-01-01" "2012-01-01" "2012-01-01" "2012-01-01" ...
 $ TSFecha : chr "2012-01-01" "2012-01-01" "2012-01-01" "2012-01-01" ...
 $ Anio : int 2012 2012 2012 2012 2012 2012 2012 2012 2012 2012 ...
 $ Mes : int 1 1 1 1 1 1 1 1 1 1 ...
 $ MesDescripcion : chr "ENE" "ENE" "ENE" "ENE" ...
 $ Dia : chr "DOMINGO" "DOMINGO" "DOMINGO" "DOMINGO" ...
 $ RepresentanteFactura : chr "MPGB" "MPGB" "MPGB" "ESAY" ...
 $ RepresentanteCliente : chr "MPGB" "MPGB" "MPGB" "ESAY" ...
 $CodigoCliente : chr "180005" "180005" "180005" "004000" ...
 $ Cliente : chr "0D5AF64D" "0D5AF64D" "0D5AF64D" "697FF5D6" ...
 $ CategorizacionCliente: chr "CLIENTE" "CLIENTE" "CLIENTE" "ALIADO" ...
 $CodigoProducto : chr "346134401015" "36134004003" "36134314003"
"346225101001" ...
 $ Producto : chr "013647BA" "07B4E1A8" "121BD315" "39CC5A4D" ...
 $ ClaseProducto : chr "COLECTIVIDAD" "COLECTIVIDAD" "COLECTIVIDAD"
"COLECTIVIDAD" ...
 $ Fabricante : chr "60A52EE9" "0092F0EF" "0092F0EF" "60A52EE9" ...
 $ GamaProducto : chr "MEDIO" "COMMODITY" "COMMODITY" "MEDIO" ...
 $ LineaProducto : chr "639EF3CC" "639EF3CC" "639EF3CC" "E99DED8F" ...
 $ TipoProducto : chr "MEDIO" "COMMODITY" "COMMODITY" "MEDIO" ...
 $ Cantidad : num -3 -6 -12 -15 -6 -15 -6 -30 -32 -30 ...
 $ Importe : num -106560 -297600 -1008000 -363660 -145464 ...
```

\$ DescuentoLinea : num -26640 -74400 -252000 -114840 -45936 ...

## Resumen del set de datos

TipoDocumento	Documento	Linea	
Length:162183	Length:162183	Min. : 1.000	
Class :character	Class :character	1st Qu.: 1.000	
Mode :character	Mode :character	Median : 2.000	
		Mean : 3.642	
		3rd Qu.: 4.000	
		Max. :149.000	
FechaId	Fecha	TSFecha	Anio
Min. :20120101	Length:162183	Length:162183	Min. :2012
1st Qu.:20130207	Class :character	Class :character	1st Qu.:2013
Median :20140610	Mode :character	Mode :character	Median :2014
Mean :20141889			Mean :2014
3rd Qu.:20160113			3rd Qu.:2016
Max. :20171229			Max. :2017
Mes	MesDescripcion	Dia	
Min. : 1.000	Length:162183	Length:162183	
1st Qu.: 4.000	Class :character	Class :character	
Median : 7.000	Mode :character	Mode :character	
Mean : 6.584			
3rd Qu.:10.000			
Max. :12.000			
RepresentanteFactura	RepresentanteCliente	CodigoCliente	
Length:162183	Length:162183	Length:162183	
Class :character	Class :character	Class :character	
Mode :character	Mode :character	Mode :character	
Cliente	CategorizacionCliente	CodigoProducto	
Length:162183	Length:162183	Length:162183	
Class :character	Class :character	Class :character	
Mode :character	Mode :character	Mode :character	
Producto	ClaseProducto	Fabricante	
Length:162183	Length:162183	Length:162183	
Class :character	Class :character	Class :character	
Mode :character	Mode :character	Mode :character	
GamaProducto	LineaProducto	TipoProducto	
Length:162183	Length:162183	Length:162183	
Class :character	Class :character	Class :character	
Mode :character	Mode :character	Mode :character	
Cantidad	Importe	DescuentoLinea	
Min. : -3790.00	Min. : -84566688	Min. : -29712620	
1st Qu.: 2.00	1st Qu.: 54570	1st Qu.: 0	
Median : 6.00	Median : 214140	Median : 6000	
Mean : 30.05	Mean : 868125	Mean : 110363	
3rd Qu.: 20.00	3rd Qu.: 680000	3rd Qu.: 64944	
Max. :10000.00	Max. :331450000	Max. : 43781280	

### Tipo de datos para las variables del set de datos

La Tabla 1, muestra la relacionan de los tipos de datos para las variables establecidas en el conjunto de datos. En general las variables son de tipo entero y caracteres, sin embargo, las variables de tipo carácter serán consolidadas como factores.

Tabla 1 *Tipo de datos para las variables del set de datos.*

<i>Nombre</i>	<i>Tipo Dato</i>
Tipo de Documento	Carácter
Documento	Carácter
Linea	Entero
Fecha Id	Entero
Fecha	Carácter
TS Fecha	Carácter
Año	Entero
Mes	Entero
Mes Descripción	Carácter
Dia	Carácter
Representante de Factura	Carácter
Representante de Cliente	Carácter
Código de Cliente	Carácter
Cliente	Carácter
Categorización de Cliente	Carácter
Código de Producto	Carácter
Producto	Carácter
Clase de Producto	Carácter
Fabricante	Carácter
Gama de Producto	Carácter
Linea de Producto	Carácter
Tipo de Producto	Carácter
Cantidad	Numérico
Importe	Numérico
Descuento de Linea	Numérico

### Vista preliminar de datos

Con el objetivo de asegurar la confidencialidad de la información, se han ofuscado las variables código y nombres de Cliente, código y descripción de productos del conjunto de datos de ventas. A continuación, se presenta vista previa de las variables de importe por

cliente, asesor comercial y productos del conjunto de datos como variables definidas para el análisis.

### **Importe de ventas diez primeros registros**

La Tabla 2, muestra un resumen de los diez primeros registros y las variables tipo de documento, documento, línea de la orden de venta, fecha de la orden, categorización del cliente y cliente del ser de datos para el análisis.

*Tabla 2 Vista previa de los diez primeros registros del set de datos*

<i>Tipo de Documento</i>	<i>Documento</i>	<i>Línea</i>	<i>Fecha</i>	<i>Categorización de Cliente</i>	<i>Cliente</i>
V20	10000913	1	40909	CLIENTE	0D5AF64D
V20	10000913	2	40909	CLIENTE	0D5AF64D
V20	10000913	3	40909	CLIENTE	0D5AF64D
V20	10000914	1	40909	ALIADO	697FF5D6
V20	10000914	2	40909	ALIADO	697FF5D6
V20	10000914	3	40909	ALIADO	697FF5D6
V20	10000914	4	40909	ALIADO	697FF5D6
V20	10000915	1	40909	CLIENTE	712A64BA
V20	10000915	2	40909	CLIENTE	712A64BA
V20	10000915	3	40,909	CLIENTE	712A64BA

### **Top 10 importe de ventas por cliente**

La Tabla 3, muestra el top 10 de ingresos de ventas por cliente y el porcentaje de participación de ventas del top 10. La descripción del cliente se ha ofuscado con el fin de asegurar la confidencialidad de la información.

Tabla 3 *Top 10 del importe total de ventas por cliente.*

<i>Cliente</i>	<i>Importe</i>	<i>% Ventas</i>
D316B628	\$20,096,512,375.00	31.59
E355568D	\$8,542,703,474.00	13.43
697FF5D6	\$6,093,708,084.00	9.58
5D58E82E	\$5,620,656,557.00	8.84
F032D838	\$4,348,609,901.00	6.84
F7EC2C3C	\$4,181,190,548.00	6.57
C88946BE	\$4,066,605,397.00	6.39
FD9874CC	\$3,767,001,428.00	5.92
D8B57E90	\$3,708,428,570.00	5.83
97C697E8	\$3,189,696,876.00	5.01

### Top 10 importe de ventas por asesor comercial

La Tabla 4, nos muestra el top 10 de ingresos de ventas por asesor comercial y el porcentaje de participación de ventas del top 10.

Tabla 4 *Top 10 del importe total de ventas por asesor comercial*

<i>Código asesor comercial</i>	<i>Importe</i>	<i>% Ventas</i>
ESAY	\$43,059,353,220.00	30.78
LIOH	\$22,954,122,082.00	16.41
SCMJ	\$20,096,512,375.00	14.37
MPGB	\$16,103,234,278.00	11.51
YMRN	\$12,900,997,159.00	9.22
AMGJ	\$9,609,226,134.00	6.87
JMSB	\$9,015,119,802.00	6.44
NYGB	\$5,030,499,303.00	3.60
GABB	\$768,054,845.00	0.55
JARU	\$360,442,152.00	0.26

### Top 10 importe de ventas por producto

La Tabla 5, nos muestra el top 10 de ingresos de ventas por producto y el porcentaje de participación de ventas del top 10.

Tabla 5 Top 10 del importe total de ventas por producto

<i>Código de Producto</i>	<i>Importe</i>	<i>% Ventas</i>
000203227050	\$3,346,933,716.00	19.16
01110000101	\$2,248,699,892.00	12.88
000203223040	\$1,841,325,830.00	10.54
002772267532	\$1,784,915,890.00	10.22
320513304001	\$1,605,412,162.00	9.19
0027722	\$1,396,665,022.00	8.00
101603214030	\$1,385,753,688.00	7.93
000204268150	\$1,329,504,821.00	7.61
001603214030	\$1,285,572,836.00	7.36
31060000001	\$1,239,037,053.00	7.09

### **Visualización básica de datos**

Se realiza la visualización básica de las variables de importe de ventas para clientes, asesor comercial, productos y descuento de línea de factura por periodo, con el fin de entender la forma en la cual los datos están distribuidos y/o representados.

La Figura 16, nos muestra el importe de ventas por fecha, para el periodo comprendido entre enero de 2012 y diciembre de 2017. Cada color representa el mes del año, siendo 1 enero, 2 febrero, 3 marzo, etc. Se evidencia importe de ventas atípicos (puntos en la parte superior e inferior), obedecen a proyectos puntuales y no a errores de registro de datos y/o facturación. Para el análisis se tendrán en cuenta los datos de ventas con y sin tratamiento de atipicidades.

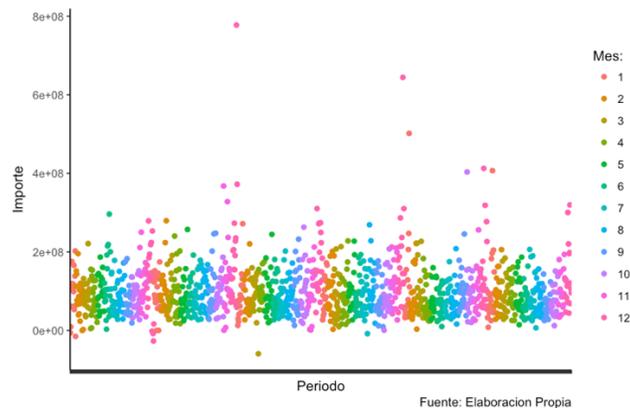


Figura 16. Importe de ventas por fecha. Elaboración propia.

La Figura 17, nos muestra el importe de ventas por año, se evidencia un incremento general del importe de ventas para el mes 12 (diciembre) y específicamente para los años 2013 y 2015. Las atipicidades obedecen a ingresos por ventas de proyectos puntuales.

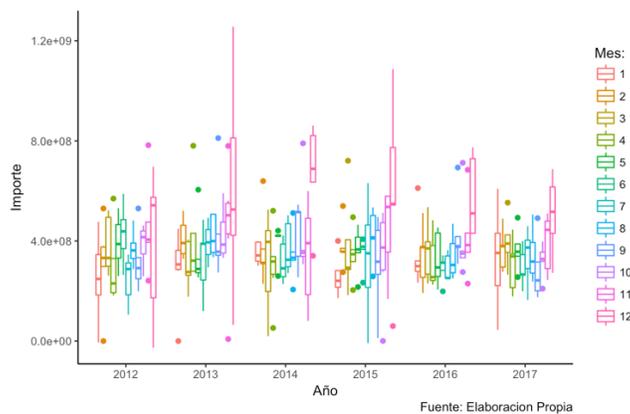


Figura 17. Importe de ventas por año. Elaboración propia.

La Figura 18, nos muestra el importe de ventas por mes y agrupados por año, se evidencia un incremento general del importe de ventas para el mes 12 (diciembre) y específicamente para los años 2013 y 2015. Las atipicidades obedecen a ingresos por ventas de proyectos puntuales.

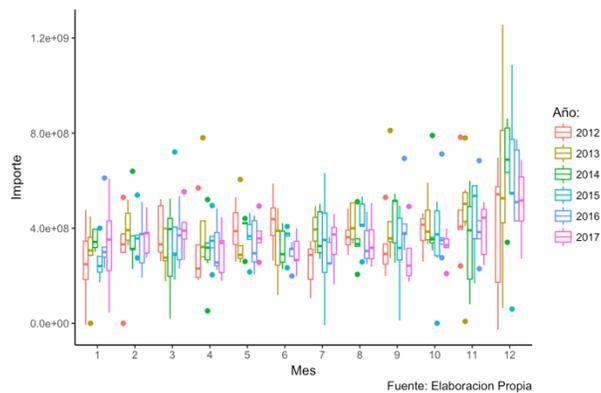


Figura 18. Importe de ventas por mes. Elaboración propia.

La Figura 19, nos muestra el importe de ventas por día de la semana, se evidencia que a excepción del domingo del año 2012 (migración de sistema ERP), no se registran ingresos por ventas, lo que al parecer implica que el domingo no se llevan a cabo operaciones. Además, el sábado representa el día con menores ingresos de importe por ventas.

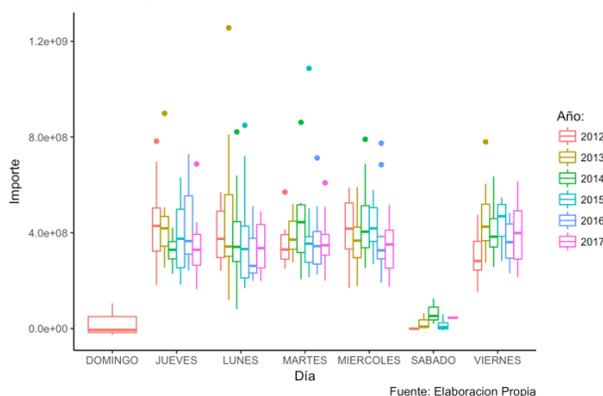


Figura 19. Importe de ventas por día de la semana. Elaboración propia.

La Figura 20, nos muestra el importe de ventas por asesor comercial, se evidencia que la mayoría de los asesores comerciales cuentan con información de histórico de ventas de al menos cinco años. Cinco de los asesores de ventas, representan mas del 50% de los ingresos de ventas. Las atipicidades obedecen a ingresos por ventas de proyectos puntuales.

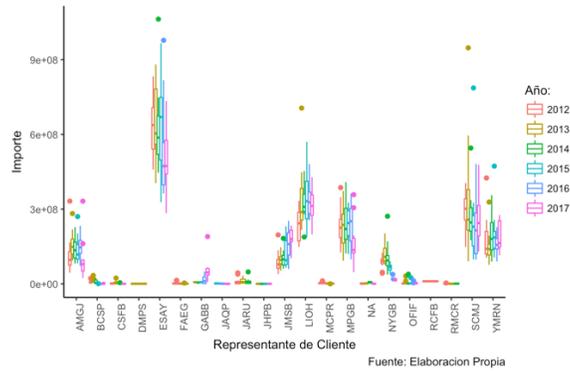


Figura 20. Importe de ventas por asesor comercial. Elaboración propia.

La Figura 21, nos muestra el importe de ventas por clase de producto, se evidencia que los productos para colectividad y oficina son los que representan el mayor importe de ventas, siendo el mes de diciembre el mes con mayores ventas. Las atipicidades obedecen a ingresos por ventas de proyectos puntuales.

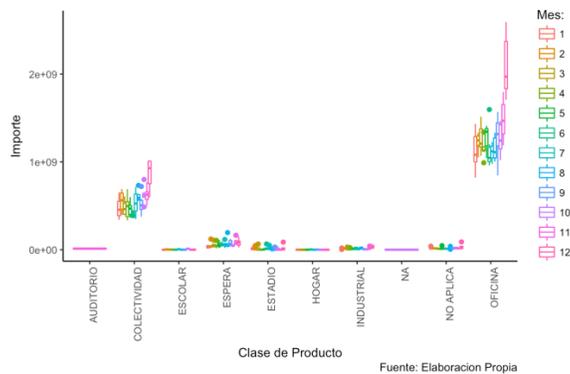


Figura 21. Importe de ventas por clase de producto. Elaboración propia.

La Figura 22, nos muestra el importe de ventas por gama de producto, se evidencia que los productos commodity y gama media, son los que representan el mayor importe de ventas. Las atipicidades obedecen a ingresos por ventas de proyectos puntuales.

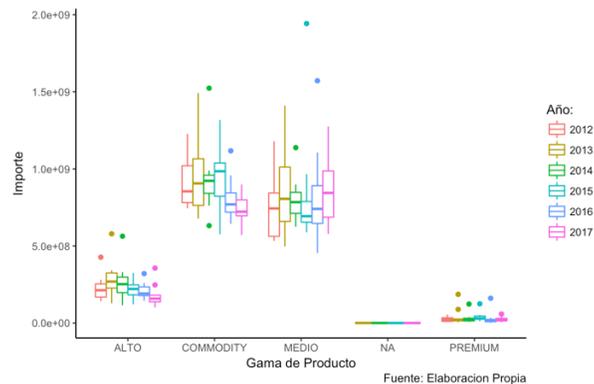


Figura 22. Importe de ventas por gama de producto. Elaboración propia.

La Figura 23, nos muestra el descuento de línea de factura por año, se evidencia que el año con mayores descuentos por línea es 2012 y se evidencia un descenso en los mismos a partir del año 2013.

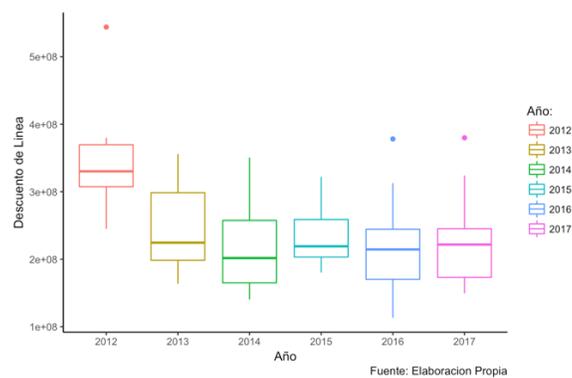


Figura 23. Descuento de línea de factura por año. Elaboración propia.

La Figura 24, nos muestra el descuento de línea de factura por mes, se evidencia que el mes de diciembre representa el mes con mayores descuentos, esto obedece quizás al incremento en el volumen de ventas.

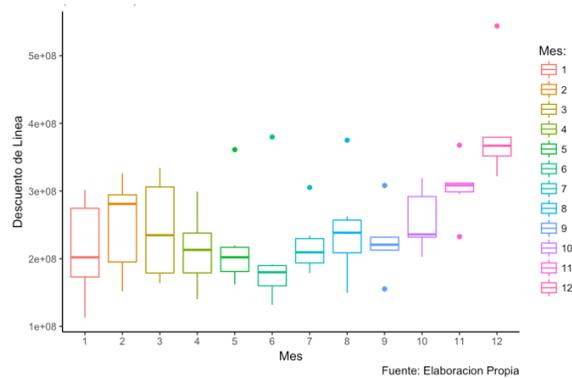


Figura 24. Descuento de línea de factura por mes. Elaboración propia.

## Preparación de datos

En la fase de preparación de datos, se contemplan las tareas de limpieza de datos, agregación de datos, generación de nuevas variables, el tratamiento de valores perdidos o faltantes y la detección de anomalías o datos atípicos.

### Tratamiento de valores perdidos (Missing Values)

Los datos para el análisis de las series de tiempo y para la generación de pronósticos, fueron tomados del repositorio de datos corporativo, cumpliendo con el proceso de transformación y normalización de datos, y los estándares de calidad definidos por el área de Informática, previo al proceso de carga.

Se utiliza el paquete Multivariate Imputation by Chained Equations (mice) de R, para la validación e imputación de valores perdidos del set de datos a analizar. La Figura 25, nos muestra gráficamente el histograma de los valores perdidos para cada una de las variables del conjunto o set de datos.

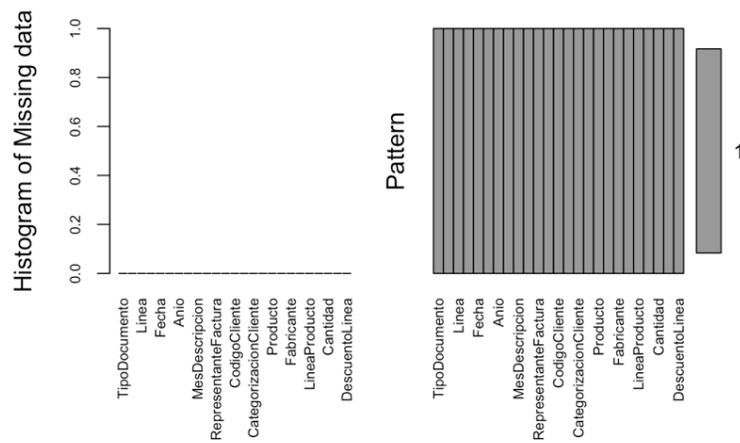


Figura 25. Histograma de valores perdidos. Elaboración propia.

La Tabla 6, nos muestra el resumen de los valores perdidos para las variables del set de datos, ordenados descendientemente por la cantidad.

Tabla 6 *Resumen de valores perdidos para las 25 variables del set de datos*

<i>Variable</i>	<i>Count</i>
Tipo de Documento	0
Documento	0
Linea	0
Fecha Id	0
Fecha	0
TS Fecha	0
Año	0
Mes	0
Mes Descripción	0
Día	0
Representante de Factura	0
Representante de Cliente	0
Código de Cliente	0
Cliente	0
Categorización de Cliente	0
Código de Producto	0
Producto	0
Clase de Producto	0
Fabricante	0
Gama de Producto	0
Linea de Producto	0
Tipo de Producto	0
Cantidad	0
Importe	0
Descuento de Linea	0

Una vez ejecutada la validación, no se evidencian valores perdidos en las variables del set de datos a procesar.

### Tratamiento de atípicos

Para la identificación de valores atípicos (observaciones con valores que no corresponden con los otros valores del set de datos), las variables contempladas en la definición de los pronósticos de ventas son el importe de ventas general, por asesor comercial y cliente y las cantidades de producto vendidos. Se realiza análisis de la distribución general y los valores atípicos, mediante el Test Grubbs y con la técnica de percentiles o Capping.

A continuación, se realiza el análisis inicial de la distribución del importe y de las cantidades de ventas general. La Figura 26, muestra la serie de datos de las ventas diarias para el periodo comprendido entre el 1 de enero de 2012 y el 31 de diciembre de 2017.

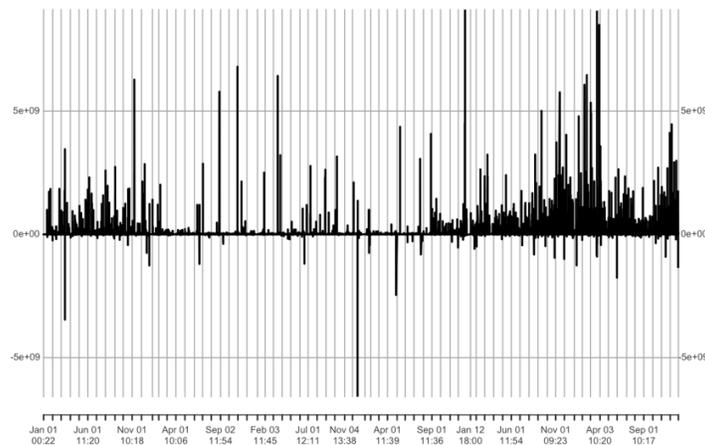
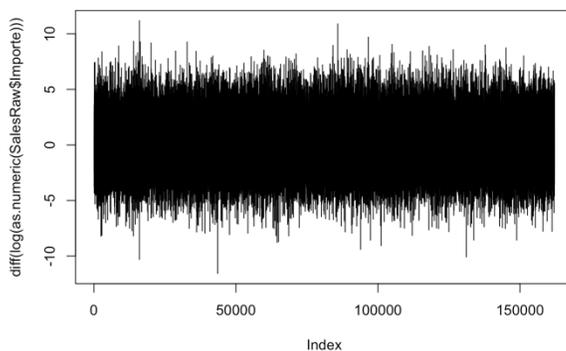


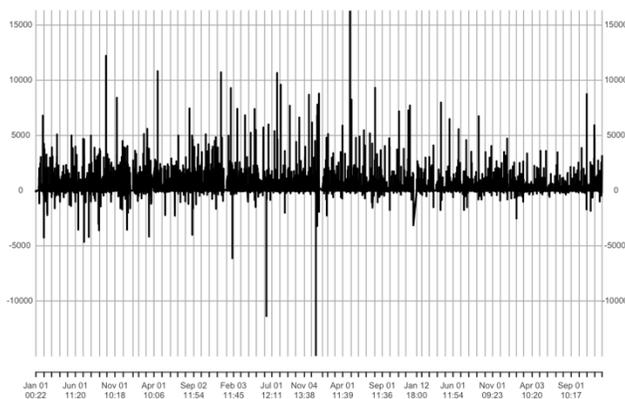
Figura 26. Serie de tiempo importe de ventas general. Elaboración propia.

La Figura 27, muestra la distribución general del importe de ventas con transformación logarítmica, esto con el fin de realizar una identificación inicial de valores atípicos.



*Figura 27.* Distribución general del importe de ventas con transformación logarítmica. Elaboración propia.

La Figura 28, muestra la serie de datos de la cantidad de productos ventas diarias para el periodo comprendido entre el 1 de enero de 2012 y el 31 de diciembre de 2017.



*Figura 28.* Serie de tiempo de venta de cantidad de productos. Elaboración propia.

La Figura 29, muestra la distribución de cantidades de productos vendidos con transformación logarítmica, esto con el fin de realizar una identificación inicial de valores atípicos

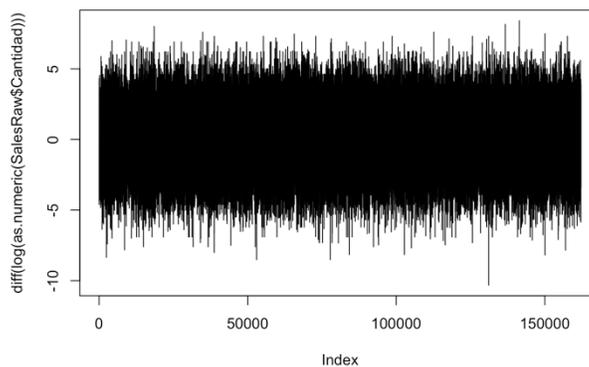


Figura 29. Distribución general de cantidades de productos vendidos con transformación logarítmica.

Elaboración propia.

La Figura 30, muestra un diagrama de caja del importe de ventas consolidado por año, con la identificación inicial del importe atípico. Se evidencian ventas con importe atípico para de los años 2012 al 2016. Se analizarán con mayor detalle con las técnicas de percentiles (Capping) y detección de anomalías (AnomalyDetection) definidas por la empresa twitter.

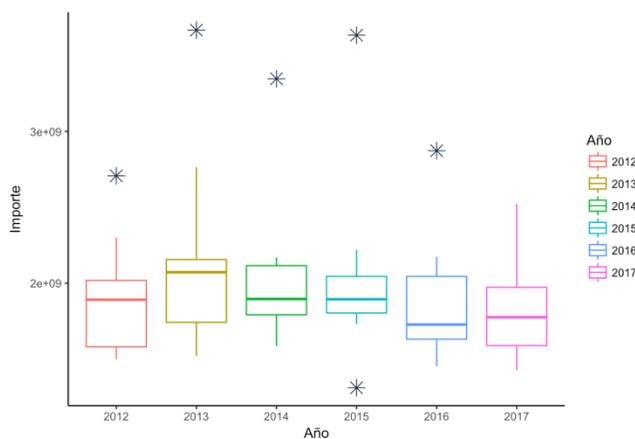


Figura 30. Importe atípico consolidado por año. Elaboración propia.

La Figura 31, muestra un diagrama de caja del importe de ventas consolidado por mes, con la identificación inicial del importe atípico. Se evidencian ventas con importe atípico para de los meses de febrero, marzo, junio, octubre y noviembre.

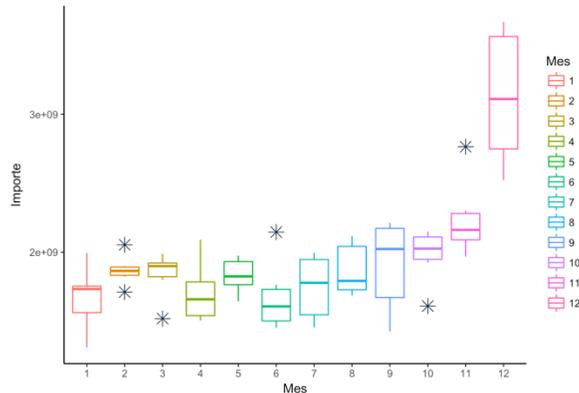


Figura 31. Importe atípico consolidado por mes. Elaboración propia.

La Figura 32, muestra un diagrama de caja de las cantidades de productos vendidos y consolidado por año, con la identificación inicial del importe atípico. Se evidencian ventas con cantidades atípicas para de los años 2014 y 2015.

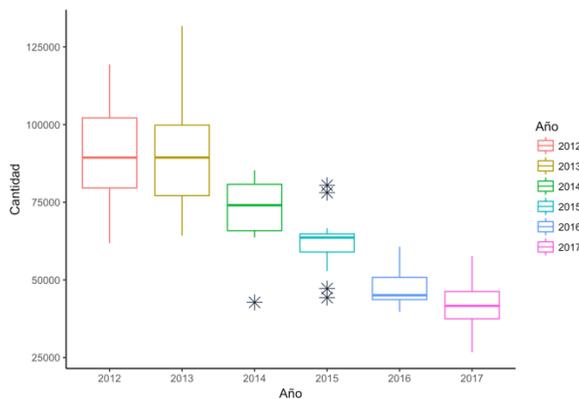


Figura 32. Cantidades atípicas consolidadas por año. Elaboración propia.

La Figura 33, muestra un diagrama de caja de las cantidades de productos vendidos y consolidado por mes, con la identificación inicial del importe atípico. No se evidencian ventas con cantidades atípicas para el periodo analizado.

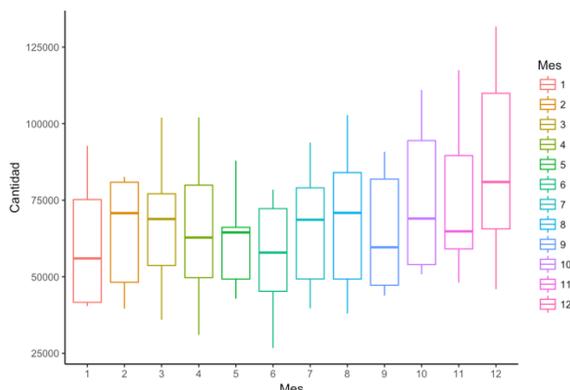


Figura 33. Cantidades atípicas consolidadas por mes. Elaboración propia.

Se realiza el test de Grubbs para determinar los valores atípicos (valores extremos máximos y mínimos opuestos) del importe de ventas general y cantidad de productos. La Figura 34, muestra el detalle del resultado del test. El resultado ofrecido por el test no es suficiente para determinar únicamente la presencia de 2 valores atípicos (valores extremos de los intervalos).

```

Grubbs test for two opposite outliers

data: SalesRaw$Importe
G = 129.27000, U = 0.93059, p-value < 2.2e-16
alternative hypothesis: -84566687.92 and 331450000 are outliers

Grubbs test for two opposite outliers

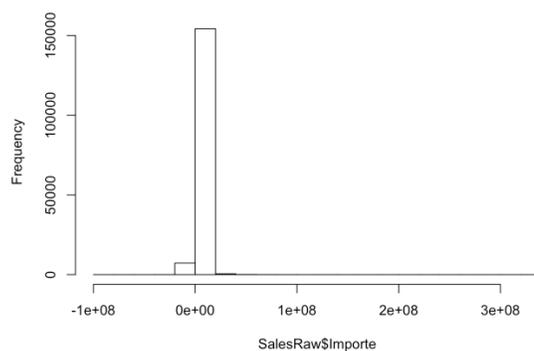
data: SalesRaw$Cantidad
G = 104.02000, U = 0.96001, p-value < 2.2e-16
alternative hypothesis: -3790 and 10000 are outliers
    
```

Figura 34. Test de Grubbs para la identificación de valores atípicos para importe y cantidad.

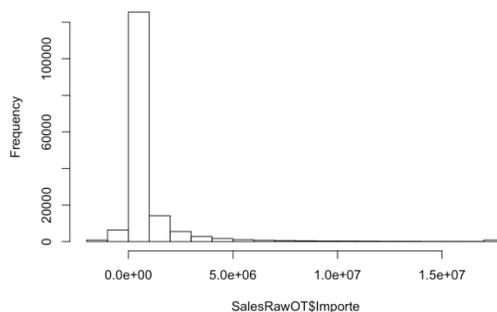
Elaboración propia.

De la validación grafica de la distribución del importe de ventas y cantidades se evidencia que existen valores mas atípicos y con el fin de identificarlos, adicionalmente aplicaremos la técnica de Capping.

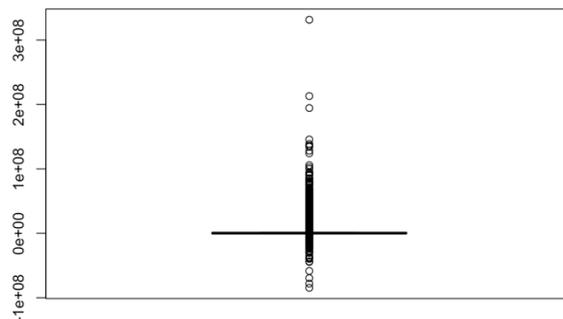
Para el tratamiento de atípicos con percentiles (Capping), se define el intervalo de confianza (Percentil Capping), en el cual los valores menores que el valor en el primer percentil son reemplazados por el valor en el primer percentil, y los valores mayores que el valor en el percentil 99 se reemplazan por el valor del percentil 99. La Figura 35, muestra un histograma del importe de ventas con un intervalo comprendido aproximadamente entre  $-1e+08$  y  $3e+08$ . Al reemplazar valores extremos con la técnica de percentiles (Capping), la Figura 36, muestra una reducción de los intervalos para el importe (consolidación de valores). En la Figura 36 y 37, podemos ver con mayor detalle la distribución del importe con y sin tratamiento de atípicos.



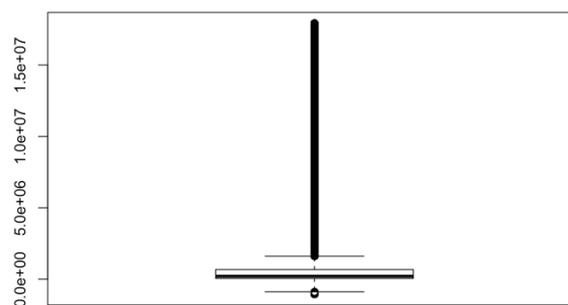
*Figura 35.* Histograma del importe de ventas. Elaboración propia.



*Figura 36.* Histograma del importe de ventas con tratamiento de atípicos. Elaboración propia.



*Figura 37.* Distribución del importe de ventas. Elaboración propia.



*Figura 38.* Distribución del importe de ventas con tratamiento de atípicos. Elaboración propia.

La Figura 39, muestra un histograma de las cantidades de productos vendidos con un intervalo comprendido aproximadamente entre -1000 y 100. Al reemplazar valores extremos con la técnica de percentiles (Capping), la Figura 40, muestra una reducción de los intervalos para las cantidades aproximadamente entre -100 y 600. En la Figura 41 y 42, podemos ver con mayor detalle la distribución de las cantidades con y sin tratamiento de atípicos.

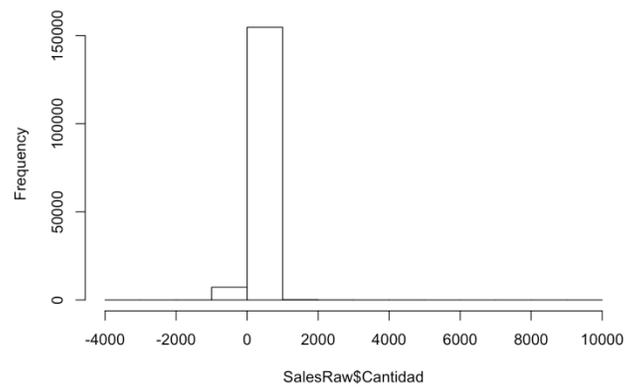


Figura 39. Histograma de cantidades. Elaboración propia.

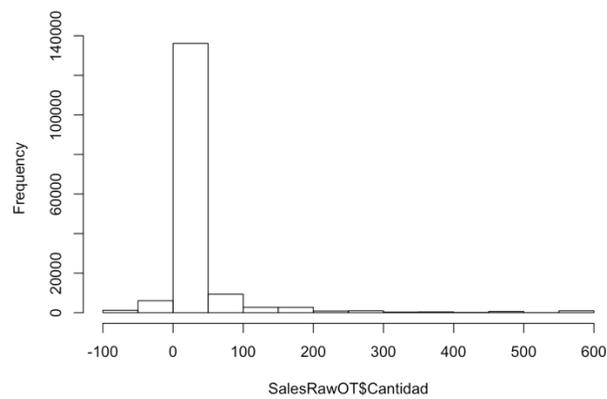


Figura 40. Histograma de cantidades con tratamiento de atípicos. Elaboración propia.

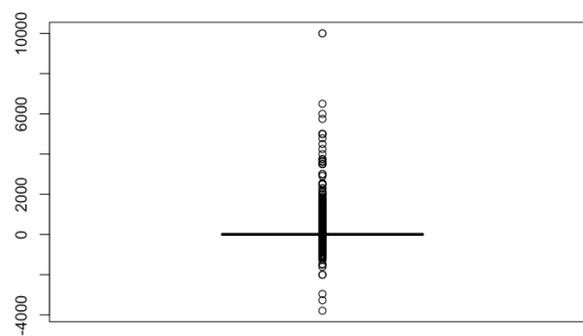


Figura 41. Distribución de cantidades. Elaboración propia.

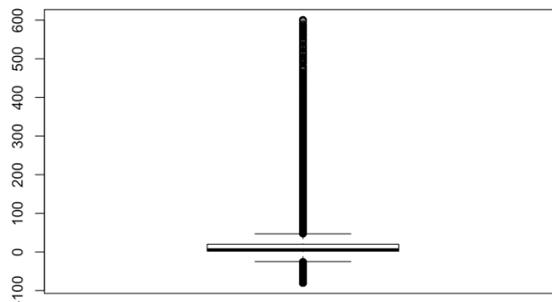


Figura 42. Distribución de cantidades con tratamiento de atípicos. Elaboración propia.

Al realizar el tratamiento de atípicos con percentiles (Capping), se observa que los datos han sido compactados (suavizados) en el intervalo establecido, adicional a una reducción en el número de observaciones de valores máximos y mínimos en importe de ventas y cantidades. Los valores negativos obedecen principalmente a devoluciones de mercancía.

Adicional realizaremos la detección de anomalías estadísticamente significativas las series de tiempo univariantes y estacionales, mediante el uso del paquete de R AnomalyDetection desarrollado por twitter. La Figura 43, muestra las anomalías detectadas para el importe de ventas, con los parámetros para la identificación del 0,2% con un  $\alpha$  de 0,05 y en ambas direcciones. La Tabla 7, muestra el consolidado de anomalías para la variable importe.

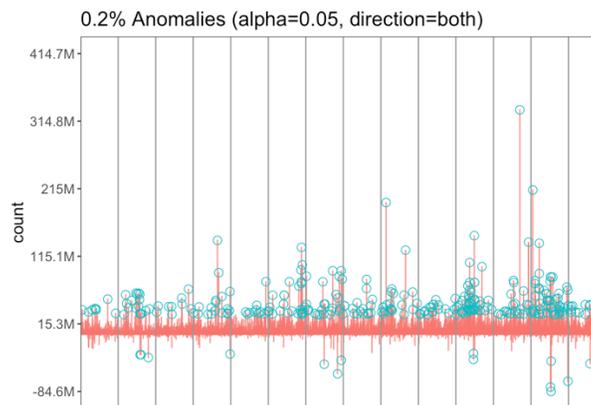


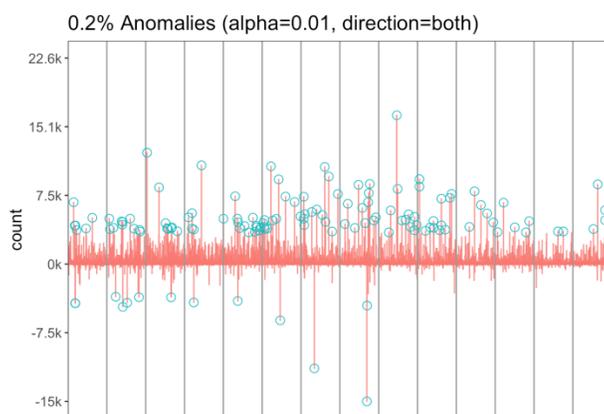
Figura 43. Detección de anomalías del importe de ventas por fecha. Elaboración propia.

Tabla 7 Consolidado de anomalías para la variable importe

<i>index</i>	<i>anoms</i>
481	35910000
2405	32832000
3606	35910000
4449	37980000
4834	37370808
4836	35890776
8422	51801120
10890	30800000
13438	28861070
14006	58231275
14291	45256320
14869	38991225
15898	49720000
17216	38562750
17217	60629212
17591	59169576
18012	41347838
18026	60629212
18319	51260400
18420	58949460

La Figura 44, muestra las anomalías detectadas para la variable cantidad, con los parámetros para la identificación del 0,2% con un alpha de 0,01 y en ambas direcciones.

La Tabla 8, muestra el consolidado de anomalías para la variable cantidad.



*Figura 44.* Detección de anomalías para la variable cantidad. Elaboración propia.

Tabla 8 *Consolidado de anomalías para la variable cantidad*

<i>index</i>	<i>anoms</i>	<i>expected_value</i>
675	6800	13
813	4248	7
854	-4248	-2
856	4248	1
885	4248	-6
886	-4248	-1
1039	3760	18
2175	3927	4
2928	5100	8
4957	5000	11
5047	3850	8
5453	4000	4
5732	-3536	8
6441	4700	31
6527	4339	1
6588	-4639	0
6591	4639	2
7118	-4197	-4
7472	4999	-2
7943	3900	43

Para la definición del presupuesto de ventas y la consolidación del plan de ventas, se presentarán los resultados de los pronósticos con y sin tratamiento de atípicos.

## **Capítulo 8**

### **Selección del modelo**

Para la fase de modelado, de acuerdo con la metodología CRISP-DM, y luego de la revisión de los enfoques implementados con éxito en la industria en la solución de problemas de regresión, se ha definido el proceso general de modelado y seleccionado los grupos y técnicas de modelado.

#### **Descripción general del proceso de modelado**

Luego de la definición de los grupos de modelos a calibrar, el proceso de modelado inicia con la carga de datos pre procesados (con tratamiento de atípicos), se realiza consolidación de ventas mensuales para ventas general, por cliente, asesor comercial y productos, de acuerdo con los criterios de priorización establecidos. Se realiza muestreo de datos en el cual el 80% corresponden a datos de entrenamiento y el 20% a datos de prueba. Se crea la serie de tiempo para el set de datos seleccionado y se realiza calibración de modelos por grupo de modelos. Se selecciona modelo del grupo, y se establecen parámetros óptimos, luego se realiza pronósticos con lo datos de entrenamiento y se mide el desempeño del modelo (proceso iterativo). El proceso termina cuando se establece la menor diferencia del pronóstico con relación al set de datos de pruebas (ver Figura 45).



Figura 45. Diagrama del proceso general de calibración de modelos. Elaboración propia.

La Figura 46, muestra el diagrama de flujo del proceso general de calibración de modelos de acuerdo con el alcance definido para el proyecto. El flujo del proceso, corresponden con cada una de las seis fases definidas.

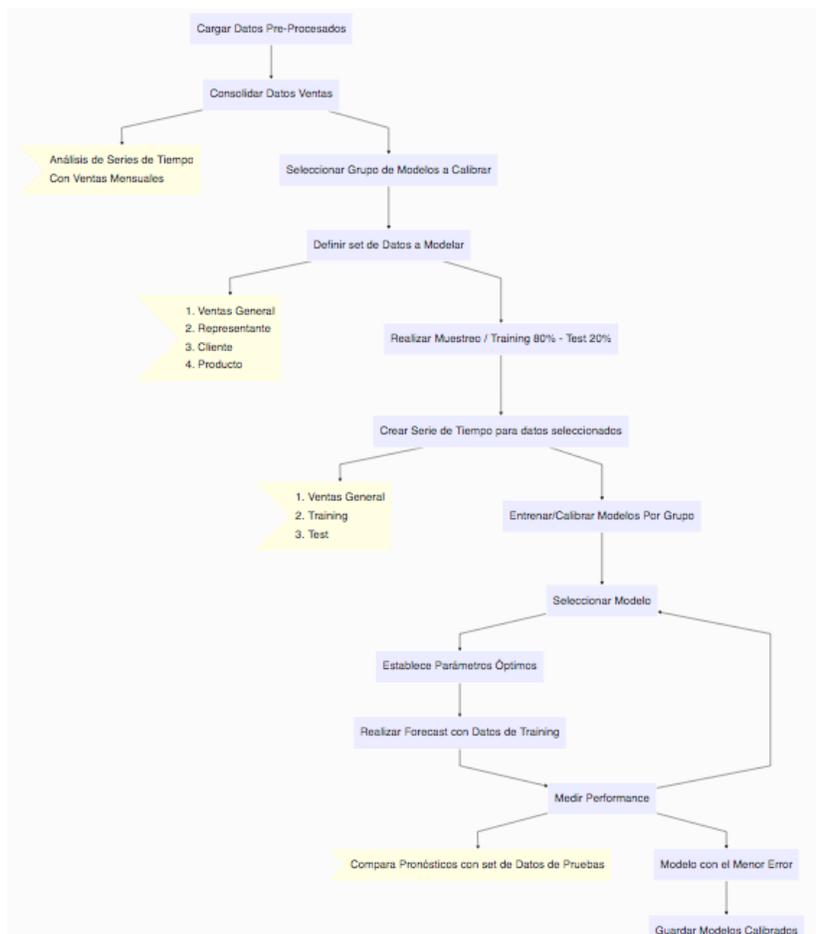


Figura 46. Flujo del proceso general de calibración de modelos. Elaboración propia.

Para la fase de modelado se deben contemplar las ventas generales, ventas por asesores comerciales, ventas por Cliente y ventas por Producto. Con el objetivo de disminuir los tiempos de procesamiento general en la definición de los pronósticos de ventas para representantes, Clientes y productos, los mismos se realizarán teniendo en cuenta únicamente los 5 modelos con el mejor desempeño. Para el pronóstico de ventas general, se analizarán los 36 modelos definidos.

## **Calibración de modelos**

### **Cargar datos pre-procesados**

Para la carga de datos se realiza consulta SQL (Structured Query Language) al Datwarehouse corporativo, con el detalle de cada una de las líneas de facturas (el numero de filas o registros es de 162183 y numero de columnas o variables es de 25). Los datos se han almacenado en un archivo con extensión csv en el directorio del proyecto con el fin de poder trabajar en caso de no tener conexión con el Datwarehouse y disminuir los tiempos de consulta. Los datos con tratamiento de atípicos fueron almacenados en un archivo de datos csv adicional. Se consolida información de datos de ventas general con tratamiento y sin tratamiento de atípicos y se almacena en un objeto R, con extensión RData. Para los datos de ventas por cliente, representante y productos, se realiza consulta SQL al archivo de datos previamente cargado en memoria con el fin de disminuir los tiempos de consulta.

### **Seleccionar grupos de modelos a calibrar**

Para el análisis y pronósticos de series de tiempo se han establecido 7 Grupos de Modelos los cuales son:

- Modelos de Regresión Lineal con y sin Estacionalidad (LR)
- Modelos Autor regresivos (AR)
- Modelos de Media Móvil Integrados Autor regresivos Estacionales (ARIMA)
- Modelos Holt-Winters (HW)
- Modelos Feed-Forward Neural Networks and Multinomial Log-Linear Models (NNet)
- Modelos Recurrent Neural Network Elman and Jordan

- Modelos Extreme Gradient Boosting (XGBoost)

### **Modelos seleccionados para el análisis**

Se han seleccionado 36 modelos distribuidos en los 7 Grupos, así:

- Modelos de Regresión Lineal con y sin Estacionalidad (LR)
  - Regresión Lineal [1]
  - Regresión Lineal con Componente de Estacionalidad [2]
- Modelos Autorregresivos (AR)
  - Autorregresivos Parámetros por Defecto [3]
  - Autorregresivos Parámetros por Defecto con Transformación Logarítmica [4]
  - Autorregresivos Parámetros Pre calculados [5]
  - Autorregresivos Parámetros Pre calculados con Transformación Logarítmica [6]
- Modelos de Media Móvil Integrados Autorregresivos Estacionales (ARIMA)
  - ARIMA con Parámetros por Defecto [7]
  - ARIMA con Transformación Logarítmica [8]
  - Auto ARIMA [9]
  - ARIMA con Parámetros Pre calculados y Peridiograma [10]
  - ARIMA con Parámetros Pre calculados, Transformación Logarítmica y Peridiograma [11]
  - ARIMA con Peridiograma [12]
  - ARIMA con Transformación Logarítmica y Peridiograma [13]
  - ARIMA con Parámetros Pre calculados [14]

- ARIMA con Parámetros Pre calculados y Transformación Logarítmica [15]
- Auto ARIMA con Stepwise [16]
- Modelos Holt-Winters (HW)
  - Holt-Winters Multiplicativo [17]
  - Holt-Winters Multiplicativo con Transformación Logarítmica [18]
  - Holt-Winters Aditivo [19]
  - Holt-Winters Aditivo con Transformación Logarítmica [20]
- Modelos Feed-Forward Neural Networks and Multinomial Log-Linear Models (NNet)
  - Artificial Neural Network Parámetros por Defecto [21]
  - Artificial Neural Network con Parámetros Pre calculados [22]
  - Artificial Neural Network con Parámetros Pre calculados y Transformación Logarítmica [23]
  - Artificial Neural Network con Parámetros Pre calculados y Datos Normalizados [24]
- Modelos Recurrent Neural Network Elman and Jordan
  - Elman Parámetros Básicos [25]
  - Jordan Parámetros Básicos [26]
- Modelos Extreme Gradient Boosting (XGBoost)
  - XGBoost Sin Tendencia y Estacionalidad Datos Simulados (Autoajustados) [27]

- XGBoost Sin Tendencia y Estacionalidad (Decompose) [28]
- XGBoost Sin Tendencia y Estacionalidad (Fourier) [29]
- XGBoost Sin Tendencia y Sin Estacionalidad [30]
- XGBoost Tendencia (Differencing) y Estacionalidad Datos Simulados (Autoajustados) [31]
- XGBoost Tendencia (Differencing) y Estacionalidad (Decompose) [32]
- XGBoost Tendencia (Differencing) y Estacionalidad (Fourier) [33]
- XGBoost Tendencia (Differencing) y Sin Estacionalidad [34]
- XGBoost Tendencia (Differencing), Estacionalidad (Decompose) y Lambda Constante [35]
- XGBoost Tendencia (Differencing), Estacionalidad (Decompose) y Lambda Pre calculada [36]

### **Realizar muestreo de datos**

Para la fase de calibración de modelos, las series de tiempo de ventas general, cliente, asesor comercial y productos se han dividido en datos de entrenamiento y pruebas. El 80% de los datos se utilizarán para el entrenamiento o training de los modelos y un 20% para medir el desempeño del modelo (80/20). Luego de calibrar los modelos con diferentes combinaciones tales como 90/10, 70/30, 60/40, los algoritmos basados en redes neuronales y Holt-Winters presentaban mejores resultados al dividir los datos en 80% de entrenamiento y 20% de prueba.

```
List of 5
$ size      : int 57
$ StarYear  : int 2012
$ StarMonth : int 1
$ EndYear   : int 2016
$ EndMonth  : int 9
```

```
List of 5
 $ size      : int 15
 $ StarYear  : int 2016
 $ StarMonth : int 10
 $ EndYear   : int 2017
 $ EndMonth  : int 12
```

### **Análisis de serie de tiempo para los datos seleccionados**

Se realiza el análisis inicial de las series de tiempo de ventas generales, Cliente, asesor comercial y producto con el mayor importe de ventas. El análisis de la serie de tiempo comprende la descomposición estacional por Loess, la auto correlación y selección de orden para los Modelos ARMA y el análisis de Peridiogramas. Se desarrollan clases y métodos en R para la definición automática de las características analizadas en las series de tiempo. La Figura 47, muestra un comparativo del importe de ventas anual, con el fin de realizar un análisis preliminar de la distribución del importe. La Figura 48, muestra la serie de tiempo de ventas, contemplando el intervalo definido por la media y la desviación estándar del importe (zona sombreada). La Figura 49, muestra la serie de tiempo de los datos de entrenamiento y la Figura 50, muestra la serie de tiempo de los datos para prueba.

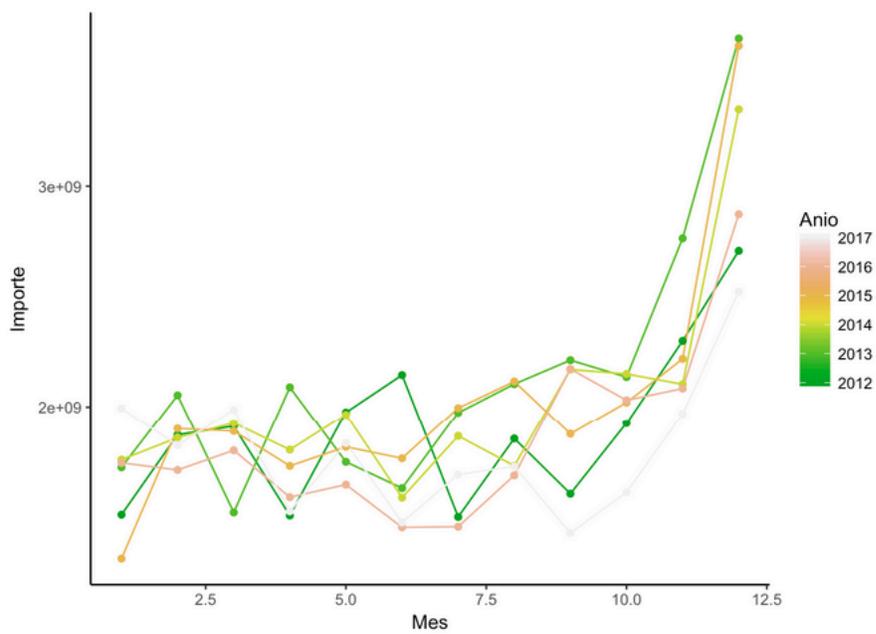


Figura 47. Comparativo del importe de ventas anual. Elaboración propia.

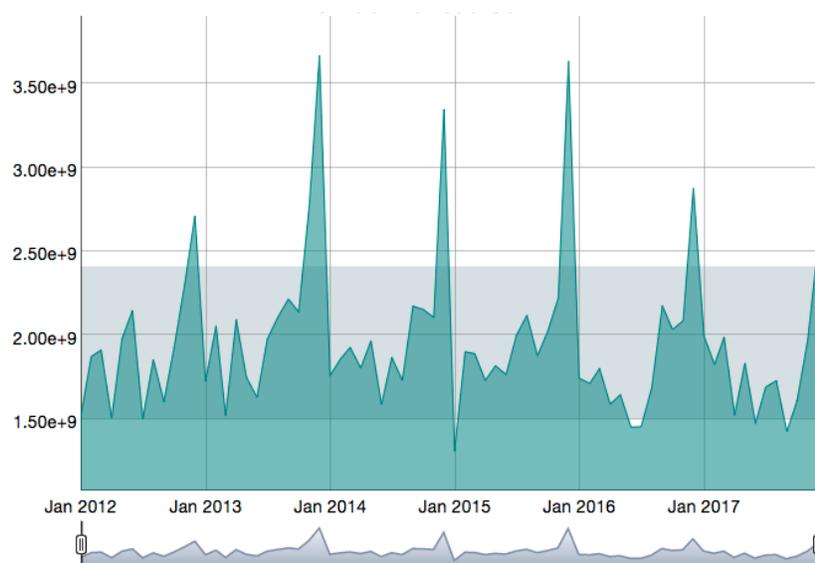


Figura 48. Serie de tiempo ventas general. Elaboración propia.

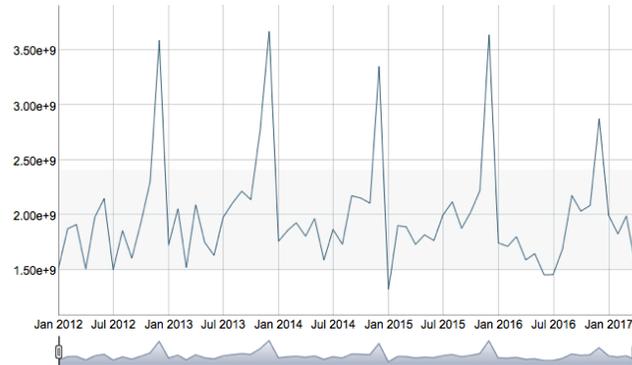


Figura 49. Serie de tiempo datos de entrenamiento ventas general. Elaboración propia.



Figura 50. Serie de tiempo datos de prueba ventas general. Elaboración propia.

La Figura 51, muestra la descomposición estacional de la serie de tiempo, con el fin de identificar gráficamente los componentes de estacionalidad, ciclicidad y tendencia. Para cada una de las series analizadas, se realiza prueba de estacionalidad.

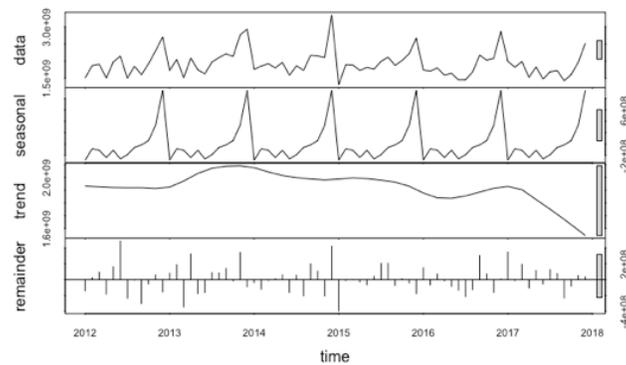


Figura 51. Descomposición estacional de la serie de tiempo ventas general por Loess. Elaboración propia.

La función de Auto Correlación (ACF) mide la correlación entre las observaciones de la serie de tiempo separadas por  $k$  períodos en el tiempo (Mensual). Y la Auto Correlación parcial mide la correlación entre dos variables separadas por  $k$  periodos cuando no se considera la dependencia creada por los retardos intermedios existentes entre ambas. La Figura 52, muestra el ACF para la serie de tiempo de entrenamiento, y de determina un ACF de grado cuatro. La Figura 53, muestra el PACF para la serie de tiempo de datos de entrenamiento, y se determina que el mismo es de grado uno. El ACF para la serie de tiempo de datos de ventas general es de grado cuatro (ver Figura 54) y PACF para la serie de tiempo de datos de ventas general es de grado dos (ver Figura 55). Estos parámetros serán utilizados para la calibración de modelos ARIMA. La Figura 56, muestra el resultado del análisis del Peridiograma de la serie de ventas general. Estos parámetros serán utilizados para la calibración de modelos ARIMA con Peridiograma.

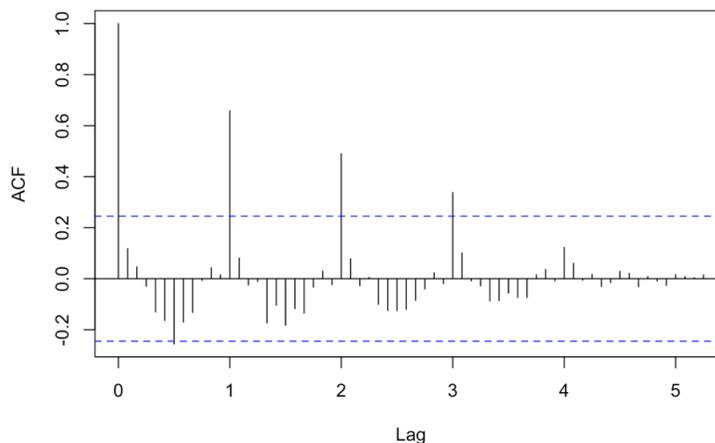


Figura 52. ACF Serie de tiempo datos de entrenamiento. Elaboración propia.

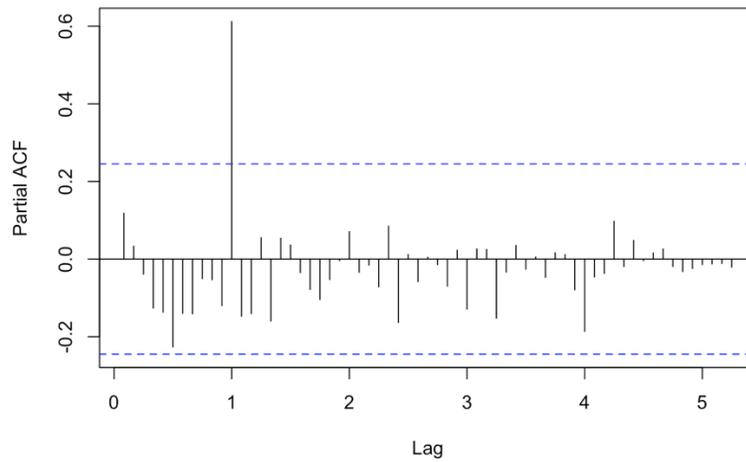


Figura 53. PACF Serie de tiempo datos de entrenamiento. Elaboración propia.

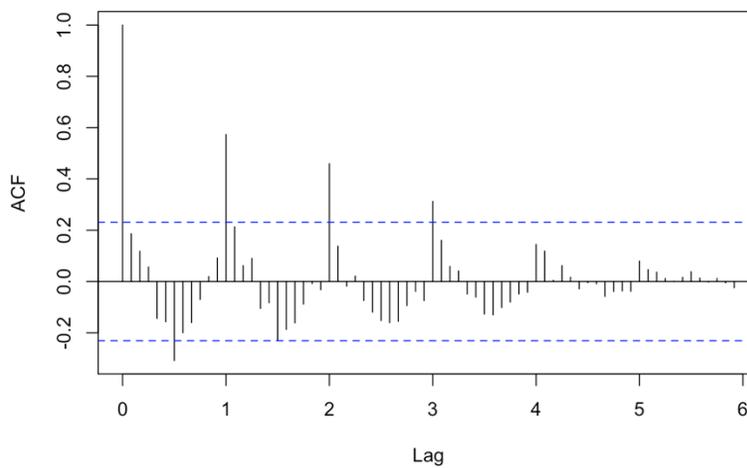


Figura 54. ACF Serie de tiempo ventas general. Elaboración propia.

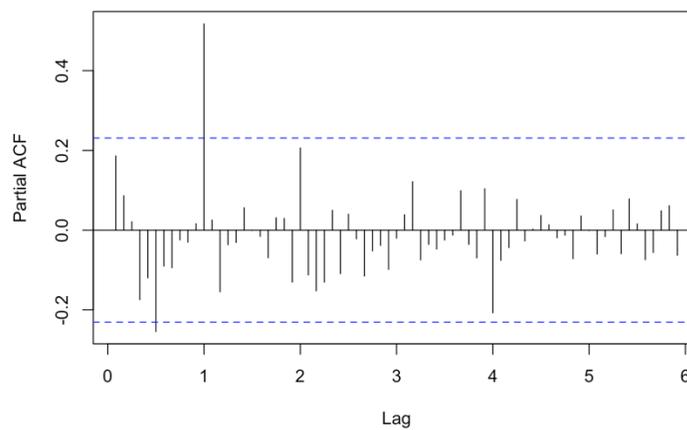


Figura 55. PACF Serie de tiempo ventas general. Elaboración propia.

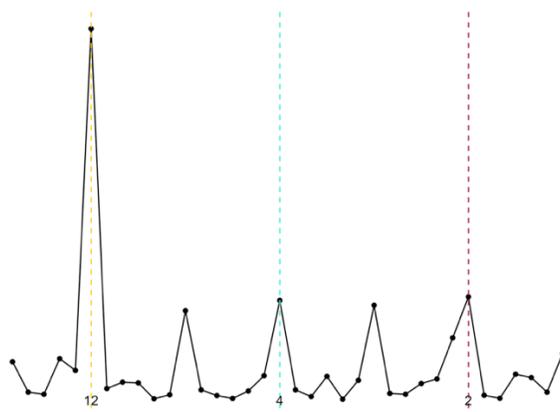


Figura 56. Peridionograma de la serie de tiempo ventas general. Elaboración propia.

Los periodos mas importantes para la serie de tiempo son el mes 12, 4 y 2 respectivamente.

### Calibrar modelos por grupo

Para la fase de calibración de modelos, y con el objetivo de optimizar el proceso de calibración de modelos para Clientes, Representantes de ventas y productos, se realiza el proceso de calibración de los 36 modelos para el set de datos de ventas general. Una vez establecidos el top 10 de modelos con el mejor desempeño para el set de datos de ventas general, estos serán utilizados para la calibración de los demás conjuntos de datos.

### Calibrar modelos de regresión lineal con y sin estacionalidad (LR)

La Tabla 9, nos muestra el resumen del tiempo de calibración en segundos de los modelos de regresión lineal con y sin estacionalidad.

Tabla 9 *Tiempo de calibración de modelos regresión lineal con y sin estacionalidad*

<i>Modelo</i>	<i>Descripción</i>	<i>Tiempo de calibración en segundos</i>
1	Regresión Lineal	0,47
2	Regresión Lineal con Componente de Estacionalidad	0,58

El tiempo de calibración del grupo de modelos regresión lineal con y sin estacionalidad (LR), es de 1.05 segundos.

### **Calibrar modelos Autorregresivos (AR)**

La Tabla 10, nos muestra el resumen del tiempo de calibración en segundos de los modelos autoregresivos.

Tabla 10 *Tiempo de calibración de modelos Autorregresivos (AR)*

<i>Modelo</i>	<i>Descripción</i>	<i>Tiempo de calibración en segundos</i>
3	Autorregresivos Parámetros por Defecto	0,25
4	Autorregresivos Parámetros por Defecto con Transformación Logarítmica	0,36
5	Autorregresivos Parámetros Pre calculados	0,59
6	Autorregresivos Parámetros Pre calculados con Transformación Logarítmica	0,44

El tiempo de calibración del grupo de modelos autorregresivo (AR), es de 1.64 segundos.

### **Calibrar modelos de media móvil integrados Autorregresivos estacionales (ARIMA)**

La Tabla 11, nos muestra el resumen del tiempo de calibración en segundos de los modelos autoregresivos estacionales o ARIMA.

Tabla 11 *Tiempo de calibración de modelos media móvil integrados*

<i>Modelo</i>	<i>Descripción</i>	<i>Tiempo de calibración en segundos</i>
7	ARIMA con Parámetros por Defecto	0,84
8	ARIMA con Transformación Logarítmica	0,41
9	Auto ARIMA	0,33
10	ARIMA con Parámetros Pre calculados y Peridiograma	16,05
11	ARIMA con Parámetros Pre calculados, Transformación Logarítmica y Peridiograma	16,49
12	ARIMA con Peridiograma	3.288,95
13	ARIMA con Transformación Logarítmica y Peridiograma	3.142,22
14	ARIMA con Parámetros Pre calculados	0,06
15	ARIMA con Parámetros Pre calculados y Transformación Logarítmica	0,06
16	Auto ARIMA con Stepwise	3,62

El tiempo de calibración del grupo de modelos media móvil integrados Autorregresivos estacionales (ARIMA) es de 6469.03 segundos.

### **Calibrar Modelos Holt-Winters (HW)**

La Tabla 12, nos muestra el resumen del tiempo de calibración en segundos de los modelos Holt-Winters.

Tabla 12 *Tiempo de calibración de modelos Holt-Winters (HW)*

<i>Modelo</i>	<i>Descripción</i>	<i>Tiempo de calibración en segundos</i>
17	HoltWinters Multiplicativo	40,08
18	HoltWinters Multiplicativo con Transformación Logarítmica	38,47
19	HoltWinters Aditivo	40,31
20	HoltWinters Aditivo con Transformación Logarítmica	38,78

El tiempo de calibración para el grupo de modelos Holt-Winters (HW), es de 157.64 segundos.

### **Calibrar Modelos Feed-Forward Neural Networks and Multinomial Log-Linear Models (NNet)**

La Tabla 13, nos muestra el resumen del tiempo de calibración en segundos de los modelos Feed-Forward Neural Networks and Multinomial Log-Linear.

Tabla 13 *Tiempo de calibración de modelos NNet*

<i>Modelo</i>	<i>Descripción</i>	<i>Tiempo de calibración en segundos</i>
21	Artificial Neural Network Parámetros por Defecto	0,72
22	Artificial Neural Network con Parámetros Pre calculados	1,50
23	Artificial Neural Network con Parámetros Pre calculados y Transformación Logarítmica	2,82
24	Artificial Neural Network con Parámetros Pre calculados y Datos Normalizados	1,24

El tiempo de calibración del grupo de modelos Feed-Forward Neural Networks and Multinomial Log-Linear Models (NNet), es de 6.28 Segundos.

### **Calibrar Modelos Recurrent Neural Network Elman and Jordan**

La Tabla 14, nos muestra el resumen del tiempo de calibración en segundos de los modelos Recurrent Neural Network Elman and Jordan.

Tabla 14 *Tiempo de Calibración de Modelos Recurrent Neural Network Elman and Jordan Neural*

<i>Modelo</i>	<i>Descripción</i>	<i>Tiempo de calibración en segundos</i>
25	Elman Parámetros Basicos	0,10
26	Jordan Parámetros Basicos	0,09

El tiempo de calibración del grupo de modelos Recurrent Neural Network Elman and Jordan Neural Network, es de 0.19 segundos.

### **Calibrar Modelos Extreme Gradient Boosting (XGBoost)**

La Tabla 15, nos muestra el resumen del tiempo de calibración en segundos de los modelos Extreme Gradient Boosting o XGBoost.

*Tabla 15 Tiempo de calibración de modelos Extreme Gradient Boosting (XGBoost)*

<i>Modelo</i>	<i>Descripción</i>	<i>Tiempo de calibración en segundos</i>
27	XGBoost Sin Tendencia y Estacionalidad Datos Simulados (Auto-Ajustados)	0,44
28	XGBoost Sin Tendencia y Estacionalidad (Decompose)	2,25
29	XGBoost Sin Tendencia y Estacionalidad (Fourier)	0,82
30	XGBoost Sin Tendencia y Sin Estacionalidad	0,44
31	XGBoost Tendencia (Differencing) y Estacionalidad Datos Simulados (Auto-Ajustados)	0,44
32	XGBoost Tendencia (Differencing) y Estacionalidad (Decompose)	0,45
33	XGBoost Tendencia (Differencing) y Estacionalidad (Fourier)	0,80
34	XGBoost Tendencia (Differencing) y Sin Estacionalidad	1,05
35	XGBoost Tendencia (Differencing), Estacionalidad (Decompose) y Lambda Constante	2,24
36	XGBoost Tendencia (Differencing), Estacionalidad (Decompose) y Lambda Pre calculada	0,49

El tiempo de calibración del grupo de modelos Extreme Gradient Boosting (XGBoost), es de 9.42 segundos.

### **Detalle del tiempo de calibración general de modelos**

La Tabla 16, nos muestra el resumen del tiempo de calibración de los modelos en segundos, y el porcentaje empleado del total del tiempo de calibración general, ordenados de menor a mayor.



Tabla 16 *Resumen de tiempo de calibración general de modelos*

<i>Modelo</i>	<i>Descripción</i>	<i>Tiempo de calibración en segundos</i>	<i>Porcentaje</i>
15	ARIMA con Parámetros Pre calculados y Transformación Logarítmica	0.06	0.0009
14	ARIMA con Parámetros Pre calculados	0.06	0.0009
26	Jordan Parámetros Básicos	0.09	0.0014
25	Elman Parámetros Básicos	0.10	0.0015
3	Autorregresivos Parámetros por Defecto	0.25	0.0038
9	Auto ARIMA	0.33	0.0050
4	Autorregresivos Parámetros por Defecto con Transformación Logarítmica	0.36	0.0054
8	ARIMA con Transformación Logarítmica	0.41	0.0062
27	XGBoost Sin Tendencia y Estacionalidad Datos Simulados (Auto-Ajustados)	0.44	0.0066
30	XGBoost Sin Tendencia y Sin Estacionalidad	0.44	0.0066
31	XGBoost Tendencia (Differencing) y Estacionalidad Datos Simulados (Auto-Ajustados)	0.44	0.0066
6	Autorregresivos Parámetros Pre calculados con Transformación Logarítmica	0.44	0.0066
32	XGBoost Tendencia (Differencing) y Estacionalidad (Decompose)	0.45	0.0068
1	Regresión Lineal	0.47	0.0071
36	XGBoost Tendencia (Differencing), Estacionalidad (Decompose) y Lambda Precalculada	0.49	0.0074
2	Regresión Lineal con Componente de Estacionalidad	0.58	0.0087
5	Autorregresivos Parámetros Pre calculados	0.59	0.0089
21	Artificial Neural Network Parámetros por Defecto	0.72	0.0108
33	XGBoost Tendencia (Differencing) y Estacionalidad (Fourier)	0.80	0.0120
29	XGBoost Sin Tendencia y Estacionalidad (Fourier)	0.82	0.0123
7	ARIMA con Parámetros por Defecto	0.84	0.0126
34	XGBoost Tendencia (Differencing) y Sin Estacionalidad	1.05	0.0158
24	Artificial Neural Network con Parámetros Pre calculados y Datos Normalizados	1.24	0.0187
22	Artificial Neural Network con Parámetros Pre calculados	1.50	0.0226

	XGBoost Tendencia (Differencing),		
35	Estacionalidad (Decompose) y Lambda Constante	2.24	0.0337
	XGBoost Sin Tendencia y Estacionalidad		
28	(Decompose)	2.25	0.0339
	Artificial Neural Network con Parámetros Pre		
23	calculados y Transformación Logarítmica	2.82	0.0424
16	Auto ARIMA con Stepwise	3.62	0.0545
	ARIMA con Parámetros Pre calculados y		
10	Peridiograma	16.05	0.2415
	ARIMA con Parámetros Pre calculados,		
11	Transformación Logarítmica y Peridiograma	16.49	0.2481
	HoltWinters Multiplicativo con Transformación		
18	Logarítmica	38.47	0.5789
	HoltWinters Aditivo con Transformación		
20	Logarítmica	38.78	0.5836
17	HoltWinters Multiplicativo	40.08	0.6031
19	HoltWinters Aditivo	40.31	0.6066
	ARIMA con Transformación Logarítmica y		
13	Peridiograma	3142.22	47.2852
12	ARIMA con Peridiograma	3288.95	49.4932

El tiempo general de calibración de los 7 grupos de modelos o 36 modelos, es de 6645.25 segundos o 110.7541667 minutos. A excepción del modelo ARIMA con transformación logarítmica y peridiograma, el proceso de calibración para los demás modelos es de menos de 60 segundos.

### **Guardar modelos calibrados**

Una vez establecidos los parámetros óptimos para cada uno de los modelos calibrados y con el objetivo de disminuir los tiempos de medición del performance y procesos de validación posteriores, se realiza el almacenamiento de cada uno de los modelos. El tiempo utilizado para el guardado de los modelos calibrados, es de 0.007 segundos.

**Medición del desempeño de los modelos o performance**

Para la definición de los parámetros óptimos en el proceso de calibración y medición del desempeño de los modelos, se realiza el pronóstico para N periodos (mismo horizonte de tiempo del set de datos de prueba) y se compara la diferencia, de acuerdo con los criterios de aceptación establecidos. Este proceso se realiza de forma iterativa hasta lograr el menor error cuadrático medio para cada uno de los modelos.

## Capítulo 9

### Medición del desempeño de los Modelos

#### Esquema general

Para la medición del desempeño y precisión de los modelos, se compara la diferencia entre los datos de prueba (información real de ventas) y el pronóstico para el periodo establecido. Para la estimación del error y realizar la clasificación del desempeño de los modelos, se utiliza el Error cuadrático medio. El error cuadrático medio (ECM) de un estimador mide el promedio de los errores al cuadrado, es decir, la diferencia entre el estimador y lo que se estima.

En esta fase se evalúan los modelos, teniendo en cuenta el cumplimiento de los criterios de éxito establecidos en el presente trabajo de grado, los cuales son: la precisión y el tiempo de calibración del modelo. La Figura 57, muestra el esquema general definido para la medición del desempeño de los modelos calibrados.



Figura 57. Esquema general proceso de medición del desempeño. Elaboración propia.

### **Carga de modelos calibrados**

En la fase de calibración de modelos, los modelos son almacenados en medios físicos (disco duro de servidor) para su posterior utilización. El proceso de medición del desempeño de los modelos inicia con la carga de dichos modelos. El tiempo empleado para la carga (de almacenamiento físico a memoria) de los 36 modelos calibrados es de 0.353 segundos.

### **Realizar pronósticos con modelos calibrados**

En esta parte del proceso, se realizan pronósticos con los datos de entrenamiento y con un horizonte de tiempo similar al grupo de datos de prueba, para luego ser comparados y establecer la diferencia.

### **Medición del desempeño de modelos calibrados**

Se compara la diferencia entre los datos de prueba (información real de ventas) y el pronóstico para cada uno de los modelos calibrados y el horizonte de tiempo establecidos.

### **Medir desempeño por grupo de modelos**

Con el fin de facilitar la medición del desempeño de los modelos, se realiza el proceso para cada uno de los modelos (36) en los siete grupos establecidos.

### **Resumen de desempeño por grupo de modelos**

Se presenta el resumen del desempeño para cada uno de los siete grupos de modelos, de forma grafica y tabla resumen de acuerdo con las métricas establecidas.

### Obtener el mejor modelo regresión lineal con y sin estacionalidad (LR)

La Figura 58, nos muestra los pronósticos de los modelos de regresión lineal con y sin estacionalidad, e incluye la grafica de los datos de prueba (ventas reales set de datos de prueba), con el fin de determinar gráficamente su desempeño.

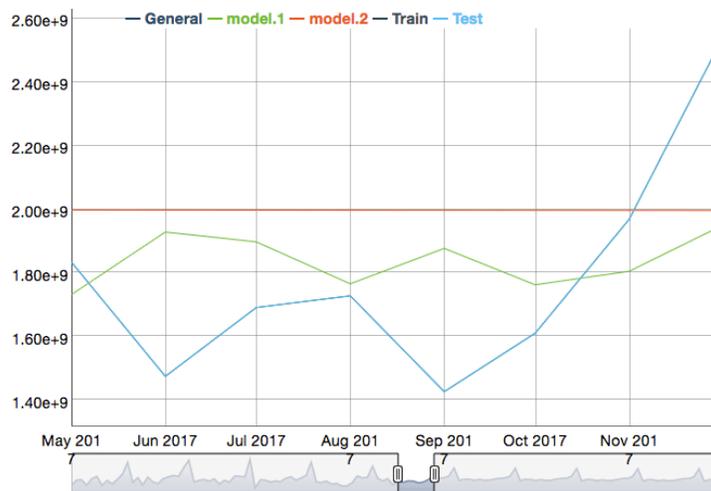


Figura 58. Pronósticos regresión lineal con y sin estacionalidad (LR). Elaboración propia.

La Tabla 17, nos muestra un resumen del desempeño de los modelos de regresión lineal con y sin estacionalidad ordenados por el error cuadrático medio (MSE) o variable objetivo-establecida para la medición del desempeño.

Tabla 17 *Desempeño de modelos de regresión lineal con y sin estacionalidad (LR)*

<i>Descripción</i>	<i>MSE</i>	<i>MAE</i>	<i>RSS</i>	<i>RMSE</i>
Regresión Lineal con Componente de Estacionalidad	1.20E+24	2.740.664.196	9.61E+24	1.095.964.440
Regresión Lineal	2.36E+24	3.162.291.864	1.89E+25	1.536.874.614

### Obtener el mejor modelo autorregresivo (AR)

La Figura 59, nos muestra los pronósticos de los modelos autorregresivos e incluye la grafica de los datos de prueba (ventas reales set de datos de prueba), con el fin de determinar gráficamente su desempeño.

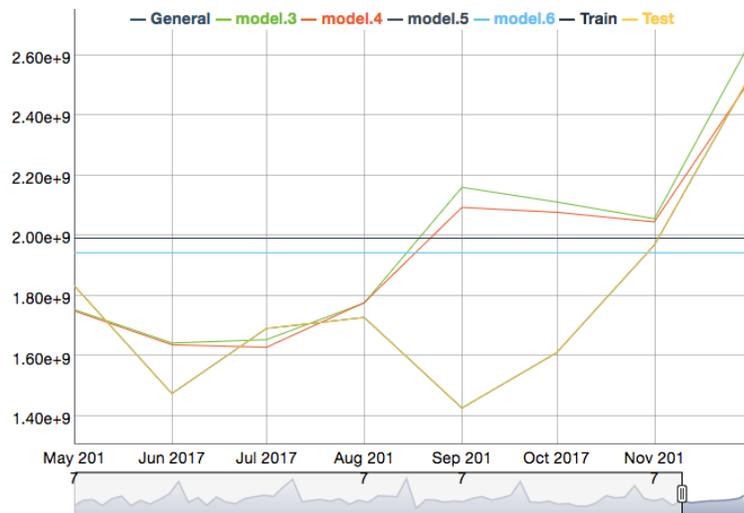


Figura 59. Pronósticos modelo autorregresivos (AR). Elaboración propia.

La Tabla 18, nos muestra un resumen del desempeño de los modelos autorregresivos ordenados por el error cuadrático medio (MSE) o variable objetivo-establecida para la medición del desempeño.

Tabla 18 *Desempeño de modelos autorregresivos (AR)*

<i>Descripción</i>	<i>MSE</i>	<i>MAE</i>	<i>RSS</i>	<i>RMSE</i>
Autorregresivos Parámetros por Defecto con Transformación Logarítmica	8.87E+22	197.782.021	7.10E+23	297.867.062
Autorregresivos Parámetros por Defecto	1.06E+23	220.978.048	8.49E+23	325.800.191
Autorregresivos Parámetros Pre calculados con Transformación Logarítmica	1.32E+23	312.465.904	1.06E+24	363.379.177
Autorregresivos Parámetros Pre calculados	1.50E+23	342.409.879	1.20E+24	387.249.087

### Obtener el mejor modelo media móvil integrados autorregresivos estacionales (ARIMA)

La Figura 60, nos muestra los pronósticos de los modelos de media móvil integrados autorregresivos estacionales, e incluye la grafica de los datos de prueba (ventas reales set de datos de prueba), con el fin de determinar gráficamente su desempeño.

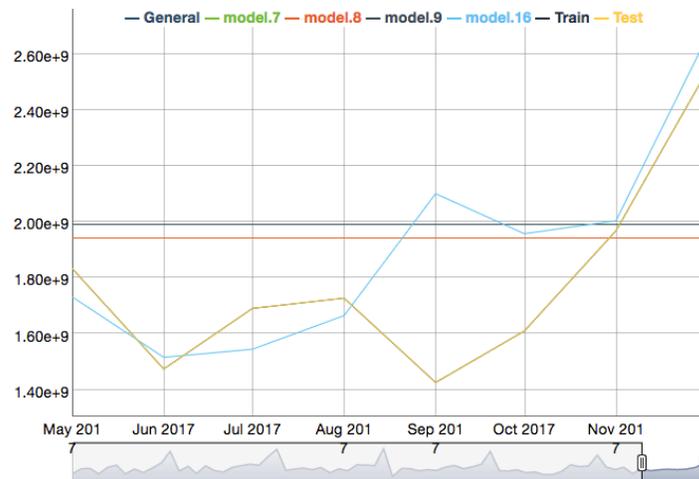


Figura 60. Pronóstico media móvil integrados autorregresivos estacionales (ARIMA). Elaboración propia.

La Tabla 19, nos muestra un resumen del desempeño de los modelos media móvil integrados autorregresivos estacionales ordenados por el error cuadrático medio (MSE) o variable objetivo-establecida para la medición del desempeño.

Tabla 19 *Desempeño de modelos de media móvil integrados autorregresivos estacionales (ARIMA)*

<i>Descripción</i>	<i>MSE</i>	<i>MAE</i>	<i>RSS</i>	<i>RMSE</i>
Auto ARIMA con Stepwise	7.87E+22	191.243.789	6.29E+23	280.490.338
ARIMA con Transformación Logarítmica	1.32E+23	312.465.904	1.06E+24	363.379.177
ARIMA con Parámetros por Defecto	1.50E+23	342.409.879	1.20E+24	387.249.087
Auto ARIMA	1.50E+23	342.409.879	1.20E+24	387.249.087
ARIMA con Peridiograma	3.03E+23	429.638.767	2.42E+24	550.349.649
ARIMA con Transformación Logarítmica y Peridiograma	3.28E+24	1.781.612.311	2.62E+25	1.811.157.165
ARIMA con Parámetros Pre calculados y Transformación Logarítmica	3.28E+24	1.781.612.311	2.62E+25	1.811.157.165
ARIMA con Parámetros Pre calculados, Transformación Logarítmica y Peridiograma	3.28E+24	1.781.612.311	2.62E+25	1.811.157.165
ARIMA con Parámetros Pre calculados	4.19E+24	1.698.125.516	3.35E+25	2.046.427.401
ARIMA con Parámetros Pre calculados y Peridiograma	4.94E+29	498.910.249.738	3.95E+30	703.099.626.571

### **Obtener el mejor modelo Holt-Winters (HW)**

La Figura 61, nos muestra las graficas de los pronósticos de los modelos Holt-Winters, e incluye la grafica de los datos de prueba (ventas reales set de datos de prueba), con el fin de determinar gráficamente su desempeño.



Figura 61. Pronóstico Holt-Winters (HW). Elaboración propia.

La Tabla 20, nos muestra un resumen del desempeño de los modelos Holt-Winters ordenados por el error cuadrático medio (MSE) o variable objetivo-establecida para la medición del desempeño.

Tabla 20 Desempeño de modelos Holt-Winters (HW)

<i>Descripción</i>	<i>MSE</i>	<i>MAE</i>	<i>RSS</i>	<i>RMSE</i>
HoltWinters Multiplicativo con Transformación Logarítmica	5.83E+22	167.850.428	4.67E+23	241.510.807
HoltWinters Aditivo con Transformación Logarítmica	5.99E+22	169.928.185	4.79E+23	244.778.633
HoltWinters Multiplicativo	6.31E+22	171.111.205	5.05E+23	251.172.757
HoltWinters Aditivo	1.04E+23	218.819.661	8.31E+23	322.284.725

**Obtener el mejor modelo Feed-Forward Neural Networks and Multinomial Log-Linear Models (NNet)**

La Figura 62, nos muestra los pronósticos de los modelos Feed-Forward Neural Networks and Multinomial Log-Linear Models, e incluye la grafica de los datos de prueba (ventas reales set de datos de prueba), con el fin de determinar gráficamente su desempeño.

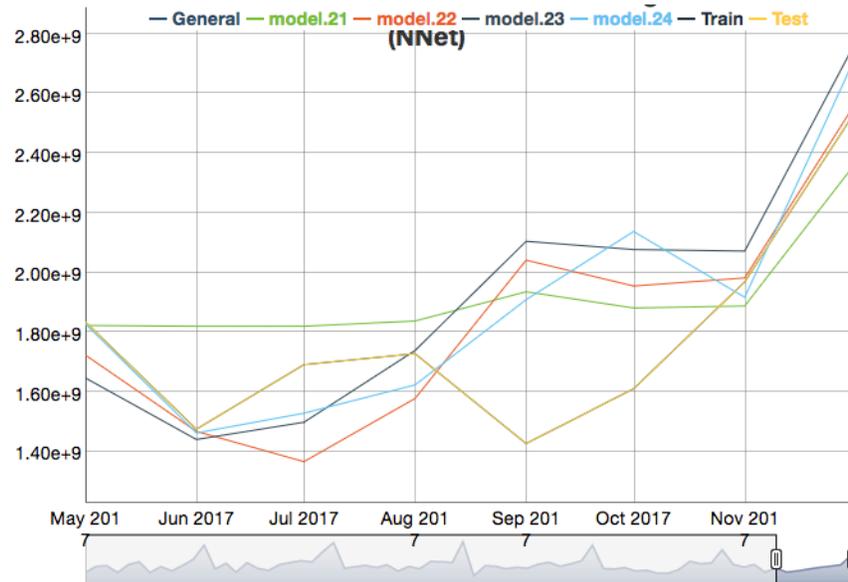


Figura 62. Pronóstico Modelos Feed-Forward Neural Networks and Multinomial Log-Linear Models

(NNet). Elaboración propia.

La Tabla 21, nos muestra un resumen del desempeño de los modelos Feed-Forward Neural Networks and Multinomial Log-Linear ordenados por el error cuadrático medio (MSE) o variable objetivo-establecida para la medición del desempeño.

Tabla 21 *Desempeño de modelos Feed-Forward Neural Networks and Multinomial Log-Linear Models*

<i>Descripción</i>	<i>MSE</i>	<i>MAE</i>	<i>RSS</i>	<i>RMSE</i>
Artificial Neural Network Parámetros por Defecto	6.45E+22	203.370.856	5.16E+23	254.036.053
Artificial Neural Network con Parámetros Pre calculados y Datos Normalizados	7.27E+22	190.364.124	5.82E+23	269.658.676
Artificial Neural Network con Parámetros Pre calculados	7.98E+22	199.175.980	6.38E+23	282.494.645
Artificial Neural Network con Parámetros Pre calculados y Transformación Logarítmica	1.02E+23	237.625.375	8.14E+23	318.931.049

### Obtener el mejor modelo Recurrent Neural Network Elman and Jordan

La Figura 63, nos muestra el comparativo del pronóstico del modelo Elman y los datos de prueba, con el fin de realizar un análisis previo del desempeño.

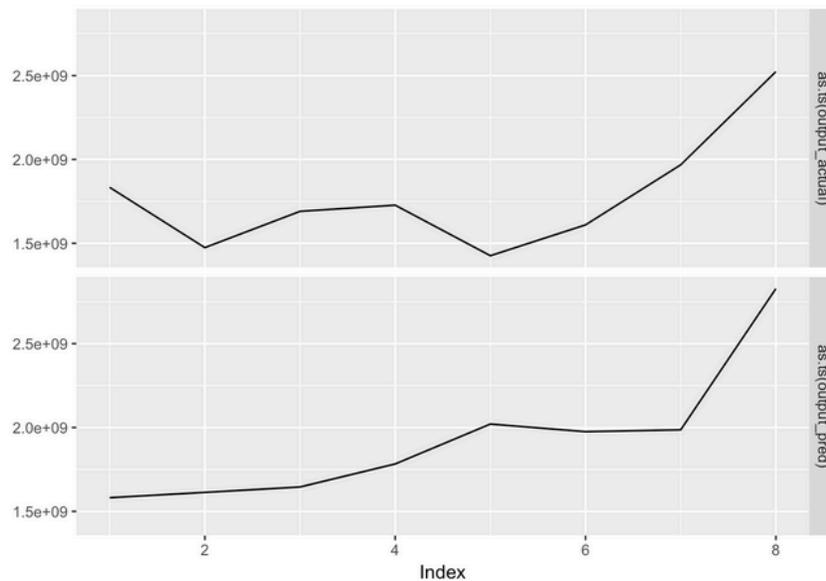


Figura 63. Pronóstico con datos de prueba modelo Elman, con parámetros básicos. Elaboración propia.

La Figura 64, nos muestra el comparativo del pronóstico del modelo Jordan y los datos de prueba, con el fin de realizar un análisis previo del desempeño.

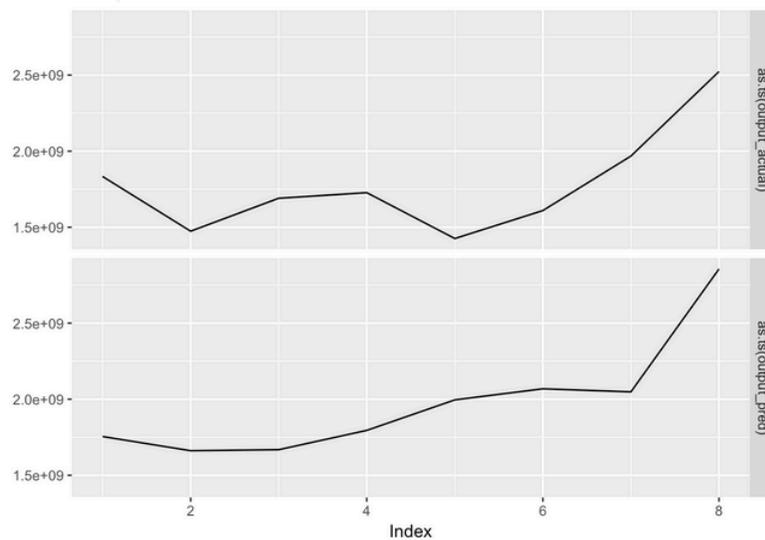


Figura 64. Pronóstico con datos de prueba modelo Jordan, con parámetros básicos. Elaboración propia.

La Tabla 22, nos muestra un resumen del desempeño de los modelos Recurrent Neural Network Elman and Jordan Neural Network ordenados por el MSE o variable objetivo-establecida para la medición del desempeño.

Tabla 22 *Desempeño de modelos Recurrent Neural Network Elman and Jordan Neural Network*

<i>Descripción</i>	<i>MSE</i>	<i>MAE</i>	<i>RSS</i>	<i>RMSE</i>
Elman Parámetros Básicos	6.44E+22	189.850.783	7.08E+23	253.762.913
Jordan Parámetros Básicos	1.07E+23	255.375.964	1.18E+24	327.146.604

### **Obtener el mejor modelo Extreme Gradient Boosting (XGBoost)**

La Figura 65, nos muestra las graficas de los pronósticos de los modelos Extreme Gradient Boosting, e incluye la grafica de los datos de prueba (ventas reales set de datos de prueba), con el fin de determinar gráficamente su desempeño.

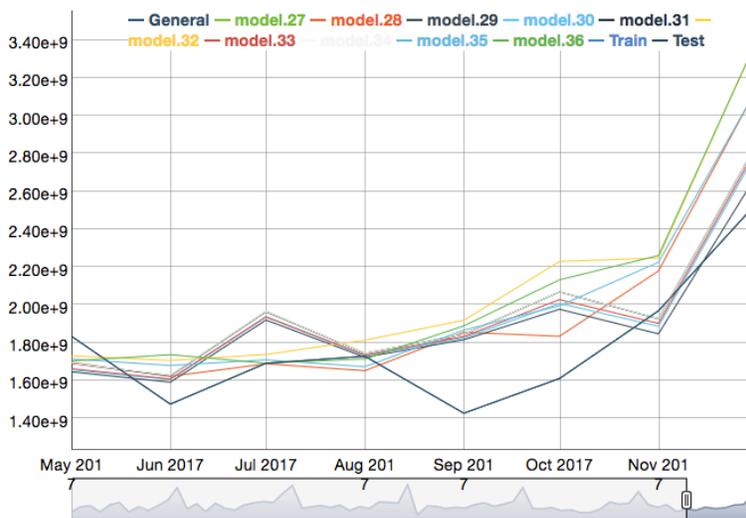


Figura 65. Pronósticos con modelos Extreme Gradient Boosting (XGBoost). Elaboración propia.

La Figura 66, nos muestra el comparativo de los 10 modelos Extreme Gradient Boosting, con el set de datos de pruebas corresponde a la línea negra y el pronostico realizado n periodos (tamaño del set de pruebas) corresponde a la línea azul, con el fin de realizar un análisis grafico del desempeño de los modelos.

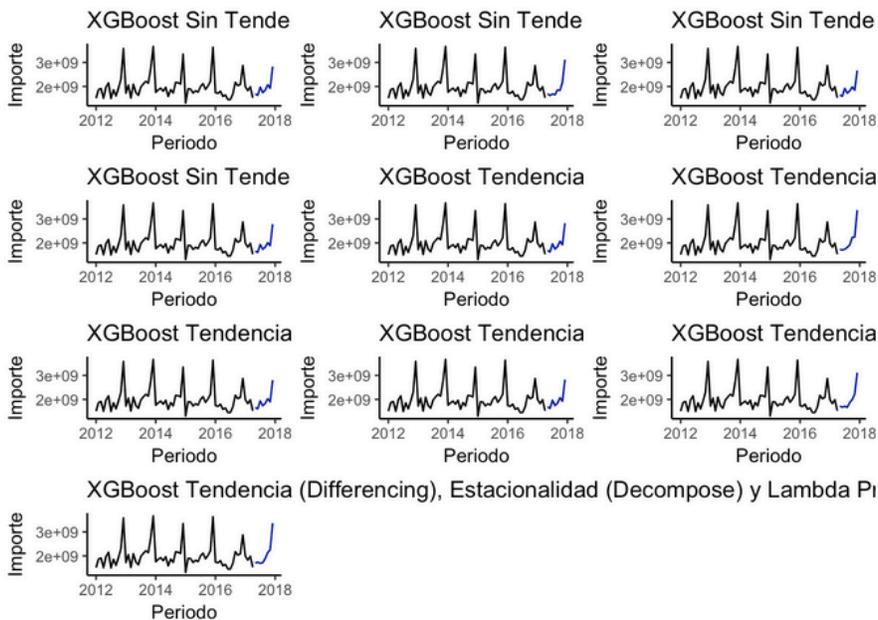


Figura 66. Comparativo de pronósticos con modelos Extreme Gradient Boosting (XGBoost).

Elaboración propia.

La Tabla 23, nos muestra un resumen del desempeño de los modelos Recurrent Neural Network Elman and Jordan Neural Network ordenados por el error cuadrático medio (MSE) o variable objetivo-establecida para la medición del desempeño.

Tabla 23 *Desempeño de modelos Extreme Gradient Boosting (XGBoost)*

<i>Description</i>	<i>MSE</i>	<i>MAE</i>	<i>RSS</i>	<i>RMSE</i>
XGBoost Sin Tendencia y Estacionalidad (Fourier)	5.27E+22	194.751.629	4.22E+23	229.631.146
XGBoost Sin Tendencia y Sin Estacionalidad	6.24E+22	211.941.869	4.99E+23	249.845.844
XGBoost Tendencia (Differencing) y Estacionalidad (Fourier)	6.62E+22	215.899.805	5.30E+23	257.346.447
XGBoost Tendencia (Differencing) y Estacionalidad Datos Simulados (Auto-Ajustados)	7.45E+22	225.301.133	5.96E+23	272.884.923
XGBoost Sin Tendencia y Estacionalidad Datos Simulados (Auto-Ajustados)	7.49E+22	225.997.453	5.99E+23	273.697.377
XGBoost Tendencia (Differencing) y Sin Estacionalidad	7.49E+22	225.997.469	5.99E+23	273.697.402
XGBoost Sin Tendencia y Estacionalidad (Decompose)	8.49E+22	227.781.140	6.79E+23	291.310.233
XGBoost Tendencia (Differencing), Estacionalidad (Decompose) y Lambda Constante	1.02E+23	259.186.501	8.20E+23	320.138.451
XGBoost Tendencia (Differencing), Estacionalidad (Decompose) y Lambda Precalculada	1.71E+23	315.969.721	1.37E+24	413.958.293
XGBoost Tendencia (Differencing) y Estacionalidad (Decompose)	1.87E+23	338.224.022	1.50E+24	432.665.697

### **Resumen de desempeño de modelos**

Una vez realizada la medición del desempeño de modelos para cada uno de los grupos, se realiza clasificación mediante la consolidación de los factores error cuadrático medio (MSE) y el tiempo de calibración de los modelos en segundos, de acuerdo con los criterios establecidos por la organización para el proyecto.

### **Resumen del desempeño de grupo de modelos con tiempo de calibración**

La Tabla 24, nos muestra el desempeño de los 7 grupos de modelos establecidos, ordenados descendientemente por el error cuadrático medio (MSE) y el tiempo de

calibración en segundos. El grupo de modelos con mejor desempeño fue Modelos Holt-Winters (HW).

Tabla 24 *Desempeño general por grupo de modelos*

<i>Grupo</i>	<i>Descripción</i>	<i>MSE</i>	<i>Tiempo de calibración en segundos</i>
4	Modelos Holt-Winters (HW)	7.13E+22	39,410
5	Modelos Feed-Forward Neural Networks and Multinomial Log-Linear Models (NNet)	7.97E+22	1,570
6	Modelos Recurrent Neural Network Elman and Jordan	8.57E+22	0,095
7	Modelos Extreme Gradient Boosting (XGBoost)	9.52E+22	0,942
2	Modelos Autorregresivos (AR)	1.19E+23	0,410
1	Modelos de Regresión Lineal con y sin Estacionalidad (LR)	1.78E+24	0,525
3	Modelos de Media Móvil Integrados Autorregresivos Estacionales (ARIMA)	4.94E+28	646903

### **Desempeño para cada uno de los modelos**

La Tabla 25, muestra el desempeño de los 36 modelos establecidos, ordenados descendientemente por el error cuadrático medio (MSE) y el tiempo de calibración en segundos.

Tabla 25 Desempeño general de modelos

<i>Modelo</i>	<i>Descripción</i>	<i>MSE</i>	<i>Tiempo de calibración en segundos</i>
29	XGBoost Sin Tendencia y Estacionalidad (Fourier)	5.27E+22	0,82
18	HoltWinters Multiplicativo con Transformación Logarítmica	5.83E+22	38,47
20	HoltWinters Aditivo con Transformación Logarítmica	5.99E+22	38,78
30	XGBoost Sin Tendencia y Sin Estacionalidad	6.24E+22	0,44
17	HoltWinters Multiplicativo	6.31E+22	40,08
25	Elman Parámetros Básicos	6.44E+22	0,10
21	Artificial Neural Network Parámetros por Defecto	6.45E+22	0,72
33	XGBoost Tendencia (Differencing) y Estacionalidad (Fourier)	6.62E+22	0,80
24	Artificial Neural Network con Parámetros Precalculados y Datos Normalizados	7.27E+22	1,24
31	XGBoost Tendencia (Differencing) y Estacionalidad Datos Simulados (Auto-Ajustados)	7.45E+22	0,44
27	XGBoost Sin Tendencia y Estacionalidad Datos Simulados (Auto-Ajustados)	7.49E+22	0,44
34	XGBoost Tendencia (Differencing) y Sin Estacionalidad	7.49E+22	1,05
16	Auto ARIMA con Stepwise	7.87E+22	3,62
22	Artificial Neural Network con Parámetros Precalculados	7.98E+22	1,50
28	XGBoost Sin Tendencia y Estacionalidad (Decompose)	8.49E+22	2,25
4	Autorregresivos Parámetros por Defecto con Transformación Logarítmica	8.87E+22	0,36
23	Artificial Neural Network con Parámetros Precalculados y Transformación Logarítmica	1.02E+23	2,82
35	XGBoost Tendencia (Differencing), Estacionalidad (Decompose) y Lambda Constante	1.02E+23	2,24
19	HoltWinters Aditivo	1.04E+23	40,31
3	Autorregresivos Parámetros por Defecto	1.06E+23	0,25
26	Jordan Parámetros Basicos	1.07E+23	0,09
6	Autorregresivos Parámetros Precalculados con Transformación Logarítmica	1.32E+23	0,44
8	ARIMA con Transformacion Logarítmica	1.32E+23	0,41
5	Autorregresivos Parámetros Precalculados	1.50E+23	0,59
7	ARIMA con Parámetros por Defecto	1.50E+23	0,84

9	Auto ARIMA	1.50E+23	0,33
	XGBoost Tendencia (Differencing), Estacionalidad (Decompose) y Lambda		
36	Precalculada	1.71E+23	0,49
	XGBoost Tendencia (Differencing) y		
32	Estacionalidad (Decompose)	1.87E+23	0,45
12	ARIMA con Peridiograma	3.03E+23	3.288,95
	Regresión Lineal con Componente de		
2	Estacionalidad	1.20E+24	0,58
1	Regresión Lineal	2.36E+24	0,47
	ARIMA con Transformación Logarítmica y		
13	Peridiograma	3.28E+24	3.142,22
	ARIMA con Parámetros Pre calculados y		
15	Transformación Logarítmica	3.28E+24	0,06
	ARIMA con Parámetros Pre calculados,		
11	Transformación Logarítmica y Peridiograma	3.28E+24	16,49
14	ARIMA con Parámetros Pre calculados	4.19E+24	0,06
	ARIMA con Parámetros Pre calculados y		
10	Peridiograma	4.94E+29	16,05

---

### Top 5 de modelos con el mejor desempeño

La Tabla 26, nos muestra el top 5 de modelos con mejor desempeño. En el encontramos dos modelos Extreme Gradient Boosting y tres modelos Holt-Winters. El tiempo de calibración de los modelos Extreme Gradient Boosting es en promedio menos de un segundo, mientras que el tiempo promedio de calibración de los modelos Holt-Winters es de 39 segundos. Para la ejecución masiva de pronósticos (Clientes / Productos), tienen prelación los modelos basados en algoritmos Extreme Gradient Boosting.

Tabla 26 *Top 5 de modelos con el mejor desempeño*

<i>Modelo</i>	<i>Descripción</i>	<i>MSE</i>	<i>Tiempo de calibración en segundos</i>
29	XGBoost Sin Tendencia y Estacionalidad (Fourier)	5.27E+22	0,82
	HoltWinters Multiplicativo con Transformación		
18	Logarítmica	5.83E+22	38,47
20	HoltWinters Aditivo con Transformación Logarítmica	5.99E+22	38,78
30	XGBoost Sin Tendencia y Sin Estacionalidad	6.24E+22	0,44
17	HoltWinters Multiplicativo	6.31E+22	40,08

### Comparativo pronósticos modelos con mejor desempeño

La Figura 67, nos muestra el comparativo de los pronósticos establecidos con los 5 modelos con el mejor desempeño y el conjunto de datos de pruebas, esto con el fin de realizar un análisis grafico previo.

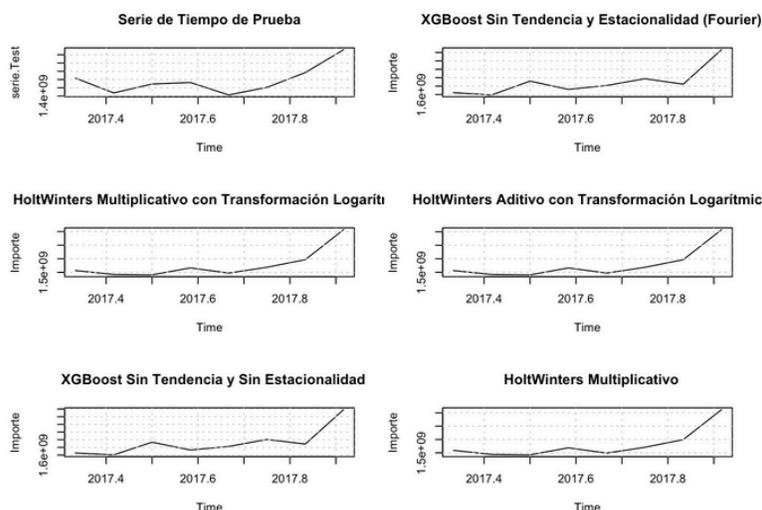


Figura 67. Comparativo de pronósticos de modelos con el mejor desempeño. Elaboración propia.

La Figura 68, nos muestra el consolidado de las series de tiempo de los pronósticos establecidos con los 5 modelos con el mejor desempeño.

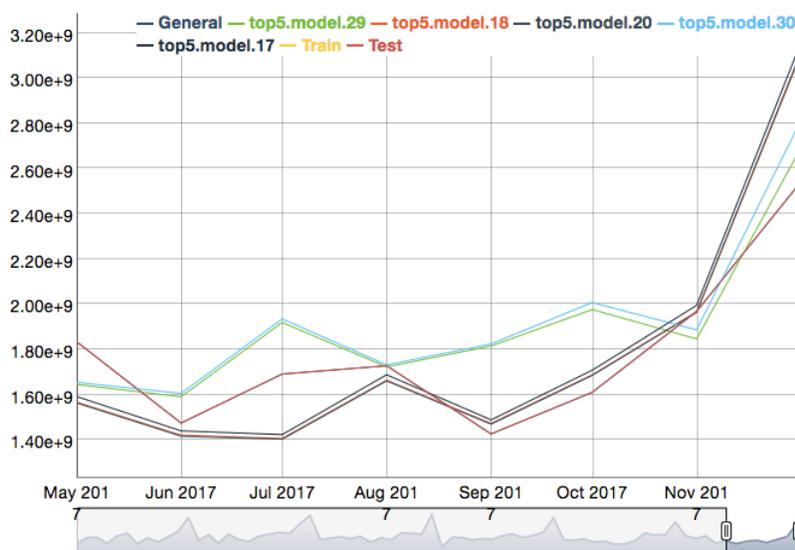


Figura 68. Comparativo de pronósticos de modelos con mejor el desempeño y consolidado.

Elaboración propia.

### Modelo con el mejor desempeño

El modelo con mejor desempeño para el set de datos de ventas general (serie de tiempo ventas general), es el Extreme Gradient Boosting XGBoost Sin Tendencia y

Estacionalidad Fourier, con un tiempo de calibración de 0,82 segundos. Se realizan cálculos iniciales para la definición de pronósticos con el set de datos completo y presenta resultados aceptables para la organización.

## Capítulo 10

### Estrategia de implementación del modelo

Para la adopción del modelo en el proceso de definición del presupuesto y el plan de ventas como parte del proceso de planeación general de la demanda, se presentará al tomador de decisiones, los pronósticos de ventas (general, asesor comercial, cliente y producto) y pronósticos con tratamiento de atípicos, la información del promedio histórico de ventas para el periodo a analizar e información de ventas del periodo en curso (ver Figura 70). Se representarán cifras y gráficas con el fin de facilitar el análisis (ve Figura 71). La Figura 69, muestra las principales fases de la estrategia de implementación del modelo.



Figura 69. Vista general de la estrategia de implementación. Elaboración propia.

SELECCION DE PARAMETROS  
Forecast

PRONOSTICOS GENERAL - GENERAL

PRONOSTICO VENTAS [2018]

GENERAL - 100% PARTICIPACION DEL TOTAL DE VENTAS 2018

Periodo	Promedio [2012-2018]	Ultimo Periodo [2018]	Forecast [2018]	Forecast Alpticos[2018]	Juicio Experto	Editar
Enero	1.610.758.293,74	1.237.547.564,00	1.992.440.000,00	1.504.100.000,00	0,00	[Icono]
Febrero	1.802.440.750,75	1.399.090.738,00	1.660.270.000,00	1.725.240.000,00	0,00	[Icono]
Marzo	1.753.703.691,52	1.244.692.925,00	1.671.420.000,00	1.655.030.000,00	0,00	[Icono]
Abril	1.706.792.175,90	-	1.645.930.000,00	1.856.210.000,00	0,00	[Icono]
Mayo	1.830.451.144,77	-	1.870.240.000,00	1.606.730.000,00	0,00	[Icono]
Junio	1.675.163.021,73	-	1.800.870.000,00	1.668.760.000,00	0,00	[Icono]
Julio	1.746.614.735,28	-	1.618.480.000,00	1.580.900.000,00	0,00	[Icono]
Agosto	1.870.020.339,80	-	1.827.240.000,00	1.939.820.000,00	0,00	[Icono]
Septiembre	1.910.577.236,23	-	1.658.250.000,00	1.902.050.000,00	0,00	[Icono]

Figura 70. Interface de consulta de pronósticos de ventas. Elaboración propia.



Figura 71. Graficas con detalle de histórico de ventas vs pronósticos ventas. Elaboración propia.

Para la definición del presupuesto de ventas con base en juicio de experto, se presentará una interface con la lista de selección de los factores de impacto tanto internos como externos definidos por el área comercial (ver Figura 72). Factores externos tales como la situación política y económica del país, cambios en la administración del estado, cambios en el mercado, si es año electoral, entre otros, y factores internos tales como el plan de Marketing definido, promociones, metas establecidas por el área Comercial, procesos de licitación en curso, entre otros, los cuales inciden positiva (suman) y negativamente (restan) sobre el valor final del presupuesto. El tomador de decisiones tendrá un campo para consignar las observaciones dadas al pronóstico y a la definición del valor

final del presupuesto, con el fin de facilitar el proceso de comunicación y aprobación del presupuesto de ventas final por parte de la dirección y del comité.

Interface de Edición de Juicio Experto

**FACTORES DE IMPACTO EXTERNOS**

-- Situación Política y Económica --	*	-- Cambios en la Administración del Estado --	*
-- Cambios en el Mercado --	*	-- Año Electoral --	*

**FACTORES DE IMPACTO INTERNOS**

-- Plan de Marketing --	*	-- Promociones --	*
-- Procesos de Licitación en Curso --	*	-- Metas Comercial --	*

**RESUMEN**

Pronóstico	Juicio Experto
1.982.440.000,00	1.982.440.000,00

**Observaciones**  
Ingrese las Observaciones del Pronóstico...

Figura 72. Interface de definición del presupuesto de ventas. Elaboración propia.

La Figura 73, muestra el proceso final de planeación demanda en la organización, luego de la adopción de la estrategia y de acuerdo con el alcance definido para el proyecto. Las cajas grises representan los nuevos componentes integrados al proceso.

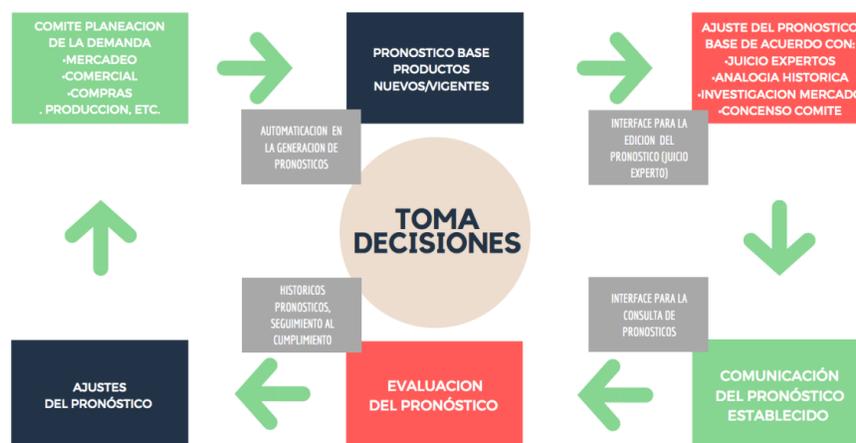


Figura 73. Proceso de planeación de la demanda propuesto. Elaboración propia.

## Capítulo 11

### Productos generados

Los productos generados y que se encuentran en ambientes productivos en la organización luego de la ejecución del proyecto, y la relación de los mismos (ver Figura 74), son:

- Interface para la Generación de Pronósticos
- Librerías para la Generación automática de Pronósticos
- Interface para la Importación Automática de Resultados de Pronósticos a la Bodega de Datos Corporativo (DWH)
- Interface para la Consulta de Pronósticos
- Interface para la Definición del Plan de Ventas, con base en Juicio de Experto
- Interface para el seguimiento al cumplimiento de Presupuesto de Ventas e
- Interface para la validación de Efectividad de Pronóstico



*Figura 74.* Vista general de los productos generados por el proyecto. Elaboración propia.

La Figura 75, describe de manera general la relación de los productos generados (entregables) con el flujo de gestión de la solución general propuesta.

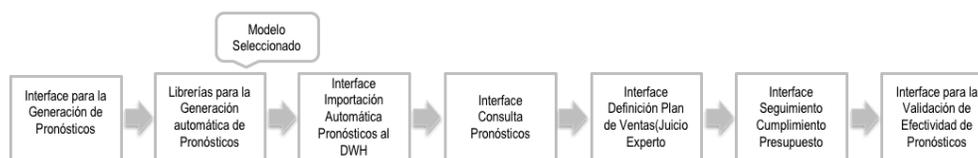


Figura 75. Productos generados por el proyecto, en orden de implementación. Elaboración propia.

## Detalle de productos generados

Luego de la definición del modelo de pronóstico de la demanda, de las pruebas de sistema y de aceptación de usuario, los productos entregados en producción a la organización son:

### Interface para la generación de pronósticos

La Figura 76, muestra la interface para la generación de pronósticos, el usuario selecciona el tipo de pronóstico y el periodo a generar, luego confirma mediante el botón Generar Pronóstico.



Figura 76. Interface para la generación de pronóstico. Elaboración propia.

### Librerías para la generación automática de pronósticos en lenguaje R

Las librerías contienen las clases y métodos para la extracción de datos desde la bodega de datos, creación y análisis de series de tiempo, calibración de modelos, carga de modelos y ejecución de pronósticos, de acuerdo con el horizonte de tiempo establecido (ver Figura 77).

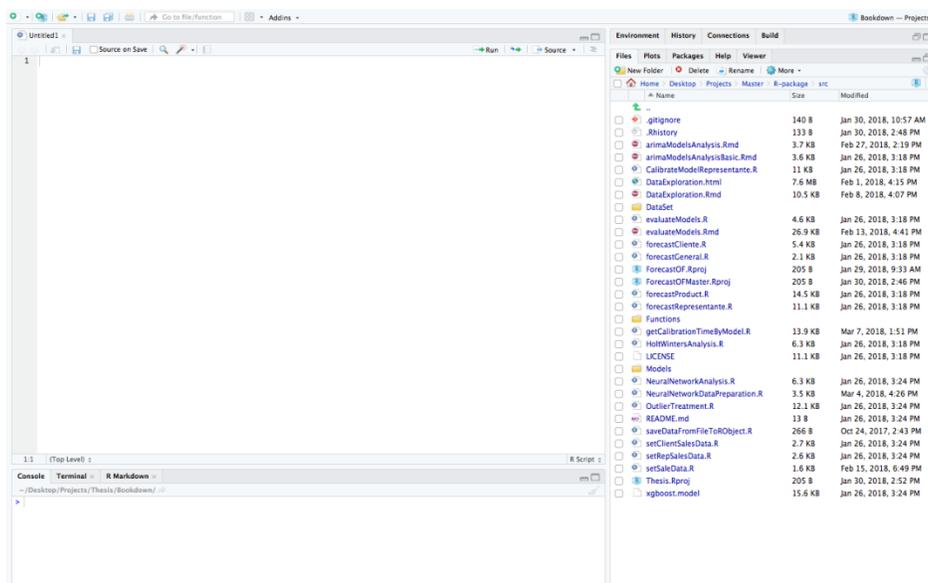


Figura 77. Estructura general de librerías del proyecto en R. Elaboración propia.

## Interface para la importación automática de resultados de pronósticos a la bodega de datos Corporativo (DWH)

Una vez ejecutado el pronóstico de acuerdo con los parámetros seleccionados, se genera un archivo en formato CSV, el cual se carga a la bodega de datos corporativa, a través de un ETL y mediante el uso de Pentaho Data Integration (Spoon). La Figura 78, muestra un el ETL de carga del archivo con pronósticos a la bodega de datos Corporativo. La Figura 79, muestra un ejemplo del archivo con los datos del pronóstico generado por el sistema, el cual será importado a la tabla de pronósticos (ver Figura 80).



Figura 78. ETL de carga de pronósticos a la bodega de datos corporativo. Elaboración propia.



Figura 79. Estructura del archivo de pronóstico generado. Elaboración propia.

ForecastGeneral_Id	GeneralForecast_Type	ForecastCode	Year	Month	Forecast	ForecastOT	Tight	ExpertJudgment	Observations	Status
128	3	200000	2.018	1	319.635.000	220.959.000	0	0	0	A
129	3	004001	2.018	1	0	0	0	0	0	A
130	3	160002	2.018	1	53.222.500	80.730.000	0	0	0	A
131	3	300000	2.018	1	88.803.600	77.928.800	0	0	0	A
132	3	400000	2.018	1	45.550.200	39.501.200	0	0	0	A
133	3	400012	2.018	1	47.400.300	63.801.300	0	0	0	A
134	3	160000	2.018	1	28.482.300	44.168.700	0	0	0	A
135	3	110083	2.018	1	112.624.000	104.525.000	0	0	0	A
136	3	000001	2.018	1	25.705.500	21.607.000	0	0	0	A
137	3	160001	2.018	1	43.143.800	31.880.500	0	0	0	A
138	3	140001	2.018	1	2.090.500	2.111.900	0	0	0	A
139	3	200020	2.018	1	29.998.400	29.998.400	0	0	0	A
140	3	120016	2.018	1	11.831.400	14.122.800	0	0	0	A
141	3	160029	2.018	1	0	0	0	0	0	A
142	3	200008	2.018	1	48.486.700	48.486.700	0	0	0	A
143	3	380000	2.018	1	12.080.900	15.793.800	0	0	0	A
144	3	002000	2.018	1	20.987.900	18.270.200	0	0	0	A
145	3	050000	2.018	1	10.315.300	10.842.900	0	0	0	A
146	3	130001	2.018	1	6.582.870	8.243.250	0	0	0	A
147	3	130022	2.018	1	6.860.480	8.218.380	0	0	0	A
148	3	200001	2.018	1	20.953.900	17.781.600	0	0	0	A
149	3	180000	2.018	1	41.832.800	26.364.700	0	0	0	A
150	3	380002	2.018	1	3.090.770	2.836.340	0	0	0	A
151	3	111112	2.018	1	7.157.770	7.157.770	0	0	0	A
152	3	130000	2.018	1	8.023.460	6.181.260	0	0	0	A
153	3	130072	2.018	1	8.882.020	8.882.020	0	0	0	A
154	3	130015	2.018	1	179.246	179.246	0	0	0	A
155	3	170000	2.018	1	21.849.300	22.637.000	0	0	0	A
156	3	120000	2.018	1	13.564.700	13.564.700	0	0	0	A
157	3	180027	2.018	1	8.572.370	10.418.400	0	0	0	A
158	3	110000	2.018	1	4.668.300	4.668.300	0	0	0	A

Figura 80. Vista de datos de pronósticos registrados en la bodega de datos. Elaboración propia.

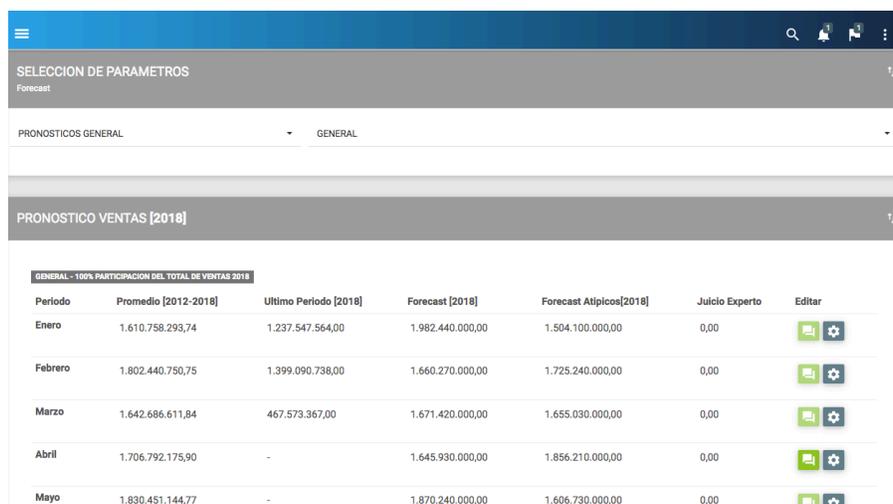
## Interface para la consulta de pronósticos

Luego del proceso de carga de los pronósticos a la bodega de datos corporativa, el usuario puede consultar los pronósticos establecidos para ventas general, asesor comercial, cliente y productos, mediante tablas y gráficos para el periodo seleccionado.

### Ventas General

La Figura 81, muestra la interface de resultado de la consulta de ventas general, la cual incluye información del periodo, promedio de ventas, pronósticos con y sin tratamiento de atípicos, presupuesto de ventas definido por el tomador de decisiones (juicio de experto). Adicional se presenta gráficamente la información de ventas y pronósticos con

y sin tratamiento de atípicos, esto con el fin de facilitar el análisis y disminuir los tiempos de toma de decisión (ver Figura 82).



Período	Promedio [2012-2018]	Último Período [2018]	Forecast [2018]	Forecast Atípicos[2018]	Juicio Experto	Editar
Enero	1.610.758.293,74	1.237.547.564,00	1.982.440.000,00	1.504.100.000,00	0,00	[Icono]
Febrero	1.802.440.750,75	1.399.090.738,00	1.660.270.000,00	1.725.240.000,00	0,00	[Icono]
Marzo	1.642.686.611,84	467.573.367,00	1.671.420.000,00	1.655.030.000,00	0,00	[Icono]
Abril	1.706.792.175,90	-	1.645.930.000,00	1.856.210.000,00	0,00	[Icono]
Mayo	1.830.451.144,77	-	1.870.240.000,00	1.606.730.000,00	0,00	[Icono]

Figura 81. Interface de consulta de pronósticos de ventas general. Elaboración propia.

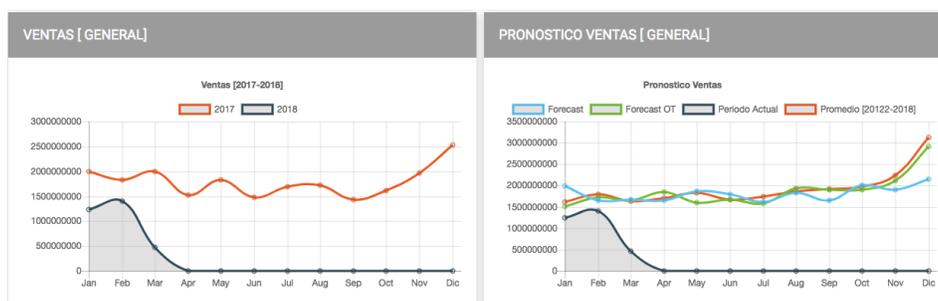


Figura 82. Interface de consulta grafica de pronósticos de ventas general. Elaboración propia.

### Ventas por Asesor Comercial

La Figura 83, muestra la interface de resultado de la consulta de ventas por asesor comercial, la cual incluye información del periodo, promedio de ventas, pronósticos con y sin tratamiento de atípicos, presupuesto de ventas definido por el tomador de decisiones (juicio de experto). Adicional se presenta gráficamente la información de ventas y pronósticos con y sin tratamiento de atípicos, esto con el fin de facilitar el análisis y disminuir los tiempos de toma de decisión (ver Figura 84).

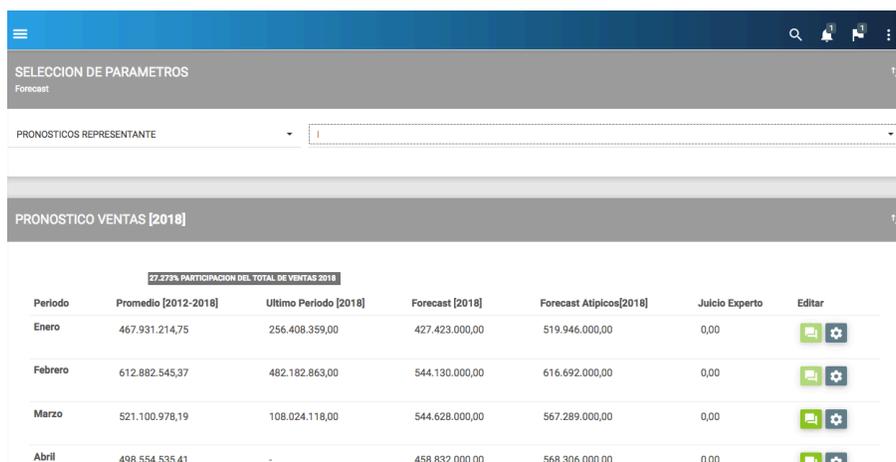


Figura 83. Interface de consulta de pronósticos por asesor comercial. Elaboración propia.

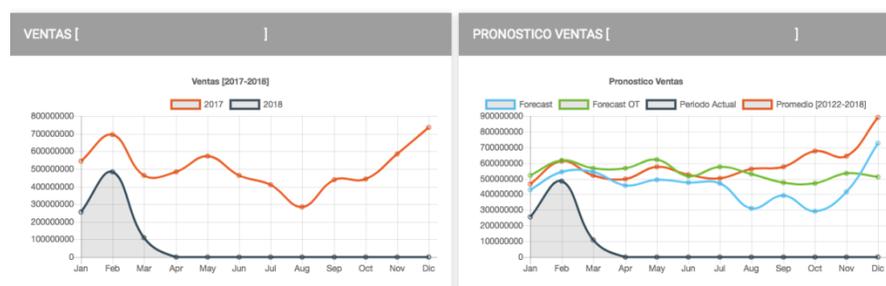


Figura 84. Interface de consulta grafica de pronósticos por asesor comercial. Elaboración propia.

## Ventas por Cliente

La Figura 85, muestra la interface de resultado de la consulta de ventas por cliente, la cual incluye información del periodo, promedio de ventas, pronósticos con y sin tratamiento de atípicos, presupuesto de ventas definido por el tomador de decisiones (juicio de experto). Adicional se presenta gráficamente la información de ventas y pronósticos con y sin tratamiento de atípicos, esto con el fin de facilitar el análisis y disminuir los tiempos de toma de decisión (ver Figura 86).

Figura 85. Interface de consulta de pronósticos por cliente. Elaboración propia.

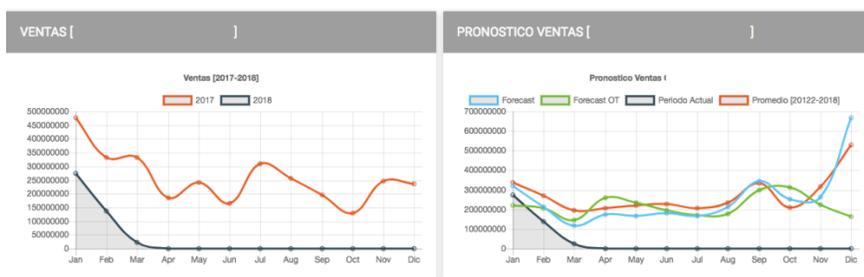


Figura 86. Interface de consulta grafica de pronósticos por cliente. Elaboración propia.

## Ventas por Producto

La Figura 87, muestra la interface de resultado de la consulta de ventas por producto, la cual incluye información del periodo, promedio de cantidad de productos vendidos, pronósticos con y sin tratamiento de atípicos, cantidades definidos por el tomador de decisiones (juicio de experto) para la consolidación del plan de ventas. Adicional se presenta gráficamente la información de ventas y pronósticos con y sin tratamiento de atípicos, esto con el fin de facilitar el análisis y disminuir los tiempos de toma de decisión (ver Figura 88).

SELECCION DE PARAMETROS  
Forecast

PRONOSTICOS PRODUCTO

PRONOSTICO VENTAS [2018]

0.697% PARTICIPACION DEL TOTAL DE VENTAS 2018

Periodo	Promedio [2012-2018]	Ultimo Periodo [2018]	Forecast [2018]	Forecast Atipicos[2018]	Juicio Experto	Editar
Enero	491,00	153,00	1.951,00	399,00	0,00	[Icono]
Febrero	454,00	260,00	885,00	332,00	0,00	[Icono]
Marzo	423,00	142,00	942,00	343,00	0,00	[Icono]
Abril	516,00	-	624,00	343,00	0,00	[Icono]

Figura 87. Interface de consulta de pronósticos por producto. Elaboración propia.

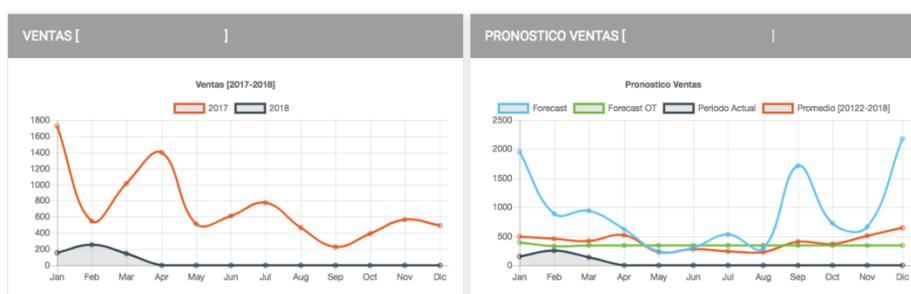


Figura 88. Interface de consulta grafica de pronósticos por producto. Elaboración propia.

### Interface para la definición del plan de ventas con base en juicio de experto

Para la definición del presupuesto de ventas y consolidación del plan de ventas, el usuario selecciona la opción definición juicio de expertos (ver Figura 89), define los factores de impacto (general o específico) (ver Figura 90), el sistema con base en los factores definidos define el presupuesto el cual puede ser modificado por el experto y finalmente ingresa las observaciones al presupuesto establecido (ver Figura 91), esto con el fin de facilitar el proceso de aprobación.

Interface de Edición de Juicio Experto

**FACTORES DE IMPACTO EXTERNOS** ⋮

-- Situación Política y Económica --	-- Cambios en la Administración del Estado --
-- Cambios en el Mercado --	-- Año Electoral --

**FACTORES DE IMPACTO INTERNOS** ⋮

-- Plan de Marketing --	-- Promociones --
-- Procesos de Licitación en Curso --	-- Metas Comercial --

**RESUMEN** ⋮

Pronóstico	Juicio Experto
1.982.440.000,00	1.982.440.000,00

**Observaciones**

Ingrese las Observaciones del Pronóstico...

---

**ACTUALIZAR**

Figura 89. Interface definición de presupuesto por juicio de experto. Elaboración propia.

Interface de Edición de Juicio Experto

**FACTORES DE IMPACTO EXTERNOS** ⋮

-- Situación Política y Económica --	-- Cambios en la Administración del Estado --
-- Cambios en el Mercado --	-- Año Electoral --

**FACTORES DE IMPACTO INTERNOS** ⋮

-- Plan de Marketing --	-- Promociones --
-- Procesos de Licitación en Curso --	-- Metas Comercial --

Figura 90. Interface de selección de factores de impacto. Elaboración propia.

RESUMEN
⋮

<b>Pronóstico</b>	<b>Juicio Experto</b>
1.982.440.000,00	1.982.440.000,00

---

**Observaciones**

Ingrese las Observaciones del Pronóstico...

ACTUALIZAR

Figura 91. Interface de resumen de juicio de experto. Elaboración propia.

### Interface para el seguimiento al cumplimiento de presupuesto

La Figura 92, muestra la interface para el seguimiento al cumplimiento del presupuesto, de acuerdo con los parámetros de periodo y asesor comercial. El sistema evalúa la información de ventas facturadas, las ordenes de venta para el periodo seleccionado pendientes por facturar, y establece el porcentaje de cumplimiento y el porcentaje de cumplimiento proyecto versus el presupuesto de ventas establecido para el periodo.

🔍 Selección de Parámetros para el Análisis de Presupuesto ⌵

Periodo: 2018 FEB FEB Representante: Todos los Representantes (10) Presupuesto Por Facturar Oportunidades

Representante	2018											
	Febrero						Total					
	Presupuesto	Facturado 2017	Facturado 2018	Por Facturar	Total	% Cumplimiento	% Proyectado	Presupuesto	Facturado 2017	Facturado 2018	Por Facturar	Total
	143.603.920,00	140.448.526,51	54.926.762,25	0,00	54.926.762,25	38,25%	38,25%	143.603.920,00	140.448.526,51	54.926.762,25	0,00	54.926.762,25
	544.130.368,00	652.754.730,44	479.185.011,45	0,00	479.185.011,45	88,06%	88,06%	544.130.368,00	652.754.730,44	479.185.011,45	0,00	479.185.011,45
	17.964.526,00	53.920.020,40	57.955.727,00	0,00	57.955.727,00	322,61%	322,61%	17.964.526,00	53.920.020,40	57.955.727,00	0,00	57.955.727,00
	210.128.800,00	172.181.389,35	173.122.081,15	0,00	173.122.081,15	82,39%	82,39%	210.128.800,00	172.181.389,35	173.122.081,15	0,00	173.122.081,15
	259.508.288,00	198.801.991,46	272.090.824,12	0,00	272.090.824,12	104,85%	104,85%	259.508.288,00	198.801.991,46	272.090.824,12	0,00	272.090.824,12
	123.477.584,00	115.353.061,90	121.078.043,12	0,00	121.078.043,12	98,06%	98,06%	123.477.584,00	115.353.061,90	121.078.043,12	0,00	121.078.043,12
	9.486.186,00	8.908.085,05	18.630.198,80	0,00	18.630.198,80	196,39%	196,39%	9.486.186,00	8.908.085,05	18.630.198,80	0,00	18.630.198,80
	2.022.426,00	1.653.335,20	8.135.572,82	0,00	8.135.572,82	402,27%	402,27%	2.022.426,00	1.653.335,20	8.135.572,82	0,00	8.135.572,82
	212.898.464,00	332.806.775,18	136.860.361,00	0,00	136.860.361,00	64,28%	64,28%	212.898.464,00	332.806.775,18	136.860.361,00	0,00	136.860.361,00
	146.664.416,00	147.085.218,12	77.106.156,08	0,00	77.106.156,08	52,57%	52,57%	146.664.416,00	147.085.218,12	77.106.156,08	0,00	77.106.156,08
<b>Grand Total</b>	<b>1.669.884.978,00</b>	<b>1.823.913.133,61</b>	<b>1.399.090.737,79</b>	<b>0,00</b>	<b>1.399.090.737,79</b>	<b>83,78%</b>	<b>83,78%</b>	<b>1.669.884.978,00</b>	<b>1.823.913.133,61</b>	<b>1.399.090.737,79</b>	<b>0,00</b>	<b>1.399.090.737,79</b>

Figura 92. Interface de seguimiento al cumplimiento del presupuesto. Elaboración propia.

## Interface para la validación de asertividad de los pronósticos

Una vez cerrado el periodo y accediendo a la información de ventas real para el periodo a analizar, se establece la diferencia en pesos y el porcentaje de variación de las ventas reales versus lo establecido por el pronóstico (ver Figura 93).

Pronóstico General						
		Evaluar Modelos	Pronóstico 2018-03	Asesor Comercial	Cliente	Producto
Salir						
> Pronóstico de Ventas						
Año	Mes	Pronóstico	Pronóstico Ajustado**	Ingresos Ventas	Diferencia (Ajustado-Ingresos)	Variación (100%)
		\$ 1.890.650.000,00	\$ 1.890.650.000,00	\$ 1.994.108.990,99	↑ \$103.458.990,99	↑ 5,19%
		\$ 1.845.910.000,00	\$ 1.845.910.000,00	\$ 1.823.913.133,61	↓ \$21.996.866,39	↓ 1,21%
		\$ 1.978.160.000,00	\$ 1.978.160.000,00	\$ 1.987.500.271,63	↑ \$9.340.271,63	↑ 0,47%
		\$ 1.851.140.000,00	\$ 1.851.140.000,00	\$ 1.524.471.442,30	↓ \$326.668.557,70	↓ 21,43%
		\$ 1.962.720.000,00	\$ 1.962.720.000,00	\$ 1.833.516.247,85	↓ \$129.203.752,15	↓ 7,05%
		\$ 1.840.260.000,00	\$ 1.840.260.000,00	\$ 1.474.235.199,66	↓ \$366.024.800,34	↓ 24,82%
		\$ 1.690.210.000,00	\$ 1.690.210.000,00	\$ 1.690.636.300,00	↑ \$426.300,00	↑ 0,03%
		\$ 1.848.160.000,00	\$ 1.848.160.000,00	\$ 1.727.458.621,00	↓ \$120.701.379,00	↓ 6,99%
		\$ 1.696.600.000,00	\$ 1.696.600.000,00	\$ 1.425.980.770,00	↓ \$270.619.230,00	↓ 16,98%
		\$ 1.878.690.000,00	\$ 1.878.690.000,00	\$ 1.609.831.517,00	↓ \$268.858.483,00	↓ 16,70%
		\$ 2.193.080.000,00	\$ 2.193.080.000,00	\$ 1.967.942.032,00	↓ \$225.137.968,00	↓ 11,44%
		\$ 2.731.830.000,00	\$ 2.594.410.000,00	\$ 2.523.297.972,00	↓ \$71.112.028,00	↓ 2,82%
		\$ 1.573.790.000,00	\$ 1.573.790.000,00	\$ 1.237.547.564,00	↓ \$336.242.436,00	↓ 27,17%
		\$ 1.660.270.000,00	\$ 1.660.270.000,00	\$ 1.399.090.736,00	↓ \$261.179.262,00	↓ 18,67%

Figura 93. Interface para la validación de la asertividad de los pronósticos. Elaboración propia.

## Capítulo 12

### Ventajas y desventajas en la adopción del modelo

#### Ventajas de la adopción del modelo

Como ventajas de la adopción del modelo de pronósticos de la demanda para la organización, tenemos:

- Automatización del proceso de generación de pronósticos de ventas general, asesor comercial, clientes y productos.
- Reducción de los tiempos de generación de pronósticos (De meses a horas)
- Mayor tiempo para el análisis de información y toma de decisión.
- Disminución del tiempo de preparación de datos para el análisis y toma de decisiones.
- Presupuestos de ventas más reales.

#### Desventajas de la adopción del modelo

Como desventajas de la adopción del modelo de pronósticos de la demanda para la organización, y de las lecciones aprendidas durante la ejecución del proyecto, tenemos:

- Requiere cambio en la Cultura Organizacional.
- Resistencia al cambio por parte de los tomadores de decisión.
- Administración de la solución general en la automatización del proceso de generación de pronósticos y administración de modelos por parte del área de tecnología.

## Capítulo 13

### Resultados obtenidos

Para la definición de los resultados, y de acuerdo con el alcance y los objetivos definidos para el proyecto, se establecieron los beneficios de la adopción del modelo para la organización, el cumplimiento de cada uno de los objetivos propuestos y el tiempo empleado en los procesos de calibración y ejecución de pronósticos con el modelo seleccionado, de acuerdo con los criterios de aceptación establecidos.

#### Beneficios para la organización

Luego de la selección del modelo, como beneficios para la organización se establecen:

- Disminución del Tiempo general de aprobación de presupuestos de ventas en un 56%.
- Disminución del tiempo de generación y procesamiento de pronósticos de ventas en un 600%.
- Disminución del tiempo de generación de información para el seguimiento al cumplimiento del presupuesto en un 96%.
- El desarrollo de capacidades de analítica predictiva en la organización permite:
  - Disminución del porcentaje de subjetividad en el proceso general de toma de decisiones.
  - Mayor eficiencia operativa.
  - Mejora en el proceso de gestión de oportunidades de ventas.

- Campañas de marketing más precisas.
- Mejora en el proceso de gestión de riesgo financiero e integración de las demás áreas en el proceso de toma de decisiones.

## Cumplimiento de los objetivos

Con base en los objetivos planteados en el presente trabajo de grado, los resultados obtenidos son:

### Caracterización de la situación actual del proceso de planificación de la demanda

Luego del análisis de la situación actual en la definición del problema de grado, se realiza el planteamiento de la solución con base en los objetivos y a las prioridades definidas por el negocio. Se establece que el alcance para el presente trabajo de grado son la automatización en la generación de pronósticos de Ventas para la definición del presupuesto y consolidación del Plan de Ventas como parte del proceso de planeación general de la demanda en la organización.

## Automatización del proceso de generación de pronósticos de ventas

### Ventas General

La Figura 94, nos muestra la definición de los pronósticos de ventas general para los 12 siguientes periodos o meses, en el repositorio de datos establecido.

ForecastGeneral_Id	GeneralForecast_Type	ForecastCode	Year	Month	Forecast	ForecastOT	ExpertJudgment	Observations
7.964	1	General	2.018	1	1.982.440.000	1.504.100.000	1.374.850.000	0
7.965	1	General	2.018	2	1.660.270.000	1.725.240.000	1.671.420.000	0
7.966	1	General	2.018	3	1.671.420.000	1.655.030.000	1.671.420.000	0
7.967	1	General	2.018	4	1.645.930.000	1.856.210.000	1.645.930.000	0 Forecast Inicial 1857690000
7.968	1	General	2.018	5	1.870.240.000	1.606.730.000	0	0
7.969	1	General	2.018	6	1.800.870.000	1.668.760.000	0	0
7.970	1	General	2.018	7	1.618.480.000	1.580.900.000	0	0
7.971	1	General	2.018	8	1.827.240.000	1.939.820.000	0	0
7.972	1	General	2.018	9	1.658.250.000	1.902.050.000	0	0
7.973	1	General	2.018	10	2.016.460.000	1.893.250.000	0	0
7.974	1	General	2.018	11	1.907.180.000	2.110.280.000	0	0
7.975	1	General	2.018	12	2.148.310.000	2.915.250.000	0	0

Figura 94. Pronósticos de venta general. Elaboración propia.

## Ventas por Asesor Comercial

Se realizan pronósticos de ventas para los asesores comerciales definidos en el alcance y con un horizonte de tiempo de 12 periodos o meses (ver Figura 95).

ForecastGeneral_Id	GeneralForecast_Type	ForecastCode	Year	Month	Forecast	ForecastOT	ExpertJudgment	Observations
7.790	2	AMGJ	2.018	1	127.409.000	127.409.000	0	0
7.784	2	ESAY	2.018	1	427.423.000	519.946.000	494.099.000	0
7.791	2	GABB	2.018	1	44.436.700	45.684.900	0	0
7.788	2	JMSB	2.018	1	171.863.000	154.909.000	0	0
7.785	2	LIOH	2.018	1	215.053.000	237.667.000	0	0
7.789	2	MPGB	2.018	1	158.222.000	325.053.000	0	0
7.792	2	NYGB	2.018	1	11.913.000	16.677.900	0	0
7.793	2	OFIF	2.018	1	1.605.860	400.174	0	0
7.786	2	SCMJ	2.018	1	319.635.000	220.959.000	0	0
7.787	2	YMRN	2.018	1	134.391.000	123.623.000	0	0

Figura 95. Pronósticos de ventas por asesor comercial. Elaboración propia.

## Ventas por Cliente

Para la definición de los pronósticos de ventas por Cliente, se realiza ley de Pareto. Se prioriza el 20% de Clientes los cuales representan el 80% de las Ventas para la organización. La Figura 96 muestra el registro de los pronósticos en el repositorio de datos definido.

ForecastGeneral_Id	GeneralForecast_Type	ForecastCode	Year	Month	Forecast	ForecastOT	ExpertJudgment	Observations
136	3	000001	2.018	1	25.705.500	21.607.000	0	0
196	3	000003	2.018	1	1.787.780	1.787.780	0	0
164	3	001000	2.018	1	11.207.100	253.430	0	0
144	3	002000	2.018	1	20.987.900	18.270.200	0	0
216	3	004000	2.018	1	87.237.100	105.914.000	0	0
145	3	050000	2.018	1	10.315.300	10.842.900	0	0
158	3	110000	2.018	1	4.668.300	4.668.300	0	0
199	3	110009	2.018	1	1.015.010	765.801	0	0
172	3	110012	2.018	1	6.717.290	6.880.600	0	0
168	3	110014	2.018	1	3.940.380	8.984.180	0	0
160	3	110015	2.018	1	3.466.870	3.218.540	0	0
214	3	110018	2.018	1	1.004.270	1.004.270	0	0
223	3	110022	2.018	1	8.467.910	8.467.910	0	0
194	3	110038	2.018	1	1.124.810	801.165	0	0
169	3	110045	2.018	1	4.088.210	5.057.520	0	0
226	3	110062	2.018	1	92.166	92.166	0	0
135	3	110083	2.018	1	112.624.000	104.525.000	0	0
206	3	110089	2.018	1	513.070	513.070	0	0
166	3	110099	2.018	1	84.379.500	84.379.500	0	0
151	3	111112	2.018	1	7.157.770	7.157.770	0	0
156	3	120000	2.018	1	13.564.700	13.564.700	0	0
167	3	120001	2.018	1	3.834.220	3.850.880	0	0
189	3	120004	2.018	1	-1.050.000	-1.050.000	0	0
201	3	120008	2.018	1	5.815.820	5.815.820	0	0
140	3	120016	2.018	1	11.831.400	14.122.800	0	0
177	3	120046	2.018	1	5.543.200	25.923.700	0	0
215	3	120049	2.018	1	7.791.020	7.791.020	0	0
152	3	130000	2.018	1	8.023.460	6.181.260	0	0
146	3	130001	2.018	1	6.582.870	8.243.250	0	0
218	3	130004	2.018	1	5.810.380	4.363.740	0	0
154	3	130015	2.018	1	179.246	179.246	0	0
162	3	130019	2.018	1	1.163.130	1.163.130	0	0
187	3	130020	2.018	1	1.819	28.586	0	0
181	3	130021	2.018	1	20.069.600	20.069.000	0	0
147	3	130022	2.018	1	6.860.480	8.218.380	0	0
191	3	130045	2.018	1	76.818.100	76.818.100	0	0
184	3	130051	2.018	1	9.839.850	9.839.850	0	0
165	3	130058	2.018	1	1.564.510	1.564.510	0	0
153	3	130072	2.018	1	8.882.020	8.882.020	0	0
183	3	130112	2.018	1	1.880.540	1.880.540	0	0
188	3	130118	2.018	1	45.862.700	45.862.700	0	0
138	3	140001	2.018	1	2.090.500	2.111.900	0	0
205	3	140014	2.018	1	3.143.320	3.143.320	0	0

Figura 96. Pronósticos de ventas por cliente. Elaboración propia.

## Ventas por Producto

Para la definición de pronósticos de ventas de productos, se realiza el análisis o segmentación ABC para la clasificación de inventarios. Se realizan pronósticos de ventas para las cantidades de productos con segmentación A y B como se muestra en la figura 97.

ForecastGeneral_Id	GeneralForecast_Type	ForecastCode	Year	Month	Forecast	ForecastOT	ExpertJudgment	Observations
1.692	4	000052299909	2.018	1	2	2	0	0
1.819	4	000203221709	2.018	1	2	2	0	0
1.802	4	000203222430	2.018	1	13	13	0	0
1.707	4	000203222440	2.018	1	9	9	0	0
1.729	4	000203222684	2.018	1	36	36	0	0
1.941	4	000203223030	2.018	1	7	7	0	0
1.657	4	00020322303125001	2.018	1	162	162	0	0
1.652	4	000203227050	2.018	1	1.951	399	0	0
1.675	4	000203227080	2.018	1	123	123	0	0
1.759	4	000204265530	2.018	1	5	5	0	0
1.670	4	000204265550	2.018	1	37	37	0	0
1.674	4	000204267030	2.018	1	72	72	0	0
1.656	4	000204267050	2.018	1	24	24	0	0
1.665	4	000204267850	2.018	1	7	7	0	0
1.667	4	000204268050	2.018	1	4	14	0	0
1.669	4	000204268150	2.018	1	8	8	0	0
1.836	4	000205267050	2.018	1	4	4	0	0
1.823	4	000205267850	2.018	1	12	12	0	0
1.843	4	000473261730	2.018	1	3	3	0	0
1.705	4	000473267830	2.018	1	51	51	0	0
1.840	4	001602311854	2.018	1	4	4	0	0
1.799	4	001602314040	2.018	1	7	7	0	0
1.663	4	001602314050	2.018	1	142	142	0	0
1.660	4	001603314030	2.018	1	3	3	0	0
1.680	4	001603315430	2.018	1	94	94	0	0
1.659	4	001603324330	2.018	1	2.841	2.841	0	0
1.758	4	002772235480	2.018	1	1	1	0	0
1.731	4	002772237580	2.018	1	1	1	0	0
1.835	4	002772267532	2.018	1	78	78	0	0
1.711	4	002773235430	2.018	1	5	5	0	0
1.887	4	002773237570	2.018	1	1	1	0	0
1.815	4	002773237580	2.018	1	16	16	0	0
1.789	4	002832264480	2.018	1	12	12	0	0
1.855	4	002832267880	2.018	1	1	1	0	0
1.809	4	0028323235450	2.018	1	49	49	0	0
1.886	4	002832327830	2.018	1	2	2	0	0
1.936	4	0028323268130	2.018	1	112	112	0	0
1.848	4	002873232050	2.018	1	25	25	0	0
1.732	4	002873234330	2.018	1	41	41	0	0
1.768	4	003692237880	2.018	1	8	8	0	0
1.816	4	003692237890	2.018	1	23	23	0	0
1.698	4	005001221340	2.018	1	94	94	0	0
1.937	4	005001221844	2.018	1	3	3	0	0

Figura 97. Pronósticos de ventas de cantidades de productos. Elaboración propia.

## Ventajas y desventajas de la adopción del modelo de pronóstico de la demanda

Las Ventajas y desventajas de la adopción del modelo de pronóstico de la demanda, se ha detallado en el Capítulo numero 13.

## Definición e implementación de la estrategia para la adopción del modelo

La definición e implementación de la estrategia para la adopción del modelo, se ha detallado en el Capítulo numero 11.

## Tiempo de calibración y ejecución de Pronósticos

A continuación, se detallan las estadísticas relacionadas con los tiempos de calibración, carga de modelos y ejecución de pronósticos, de acuerdo con el número de set de datos y modelos procesados, para:

### Ventas General

La Tabla 27, nos muestra un resumen de los tiempos de calibración, carga de modelos y ejecución de pronósticos para el set de datos de ventas general, de acuerdo con el alcance definido para el proyecto.

*Tabla 27. Estadísticas generales de ejecución de pronósticos de venta general*

<i>Concepto</i>	<i>Descripción</i>
Tipo	Ventas General
Número de Conjunto de Datos (Set)	1
Número de Modelos Calibrados	36
Tiempo Calibración (Segundos) por set	6643,48
Tiempo Carga de Modelos (Segundos)	0,239
Tiempo Calculo Pronóstico (Segundos) / set	1,52
Tiempo Total (Segundos) por set	6645.25
Tiempo Total (Segundos)	6645.25
Tiempo Total (Minutos)	110.75
Tiempo Total (Horas)	1,85

### **Ventas por asesor comercial**

La Tabla 28, nos muestra un resumen de los tiempos de calibración, carga de modelos y ejecución de pronósticos para el set de datos de ventas para los 12 asesores comerciales.

*Tabla 28. Estadísticas generales de ejecución de pronósticos de ventas por asesor comercial*

<i>Concepto</i>	<i>Descripción</i>
Tipo	Ventas General
Número de Conjunto de Datos (Set)	1
Número de Modelos Calibrados	36
Tiempo Calibración (Segundos) por set	6643,48
Tiempo Carga de Modelos (Segundos)	0,239
Tiempo Calculo Pronóstico (Segundos) / set	1,52
Tiempo Total (Segundos) por set	6645,25
Tiempo Total (Segundos)	6645,25
Tiempo Total (Minutos)	110,75
Tiempo Total (Horas)	1,85

### **Ventas por Cliente**

La Tabla 29, nos muestra un resumen de los tiempos de calibración, carga de modelos y ejecución de pronósticos para el set de datos de ventas para los clientes priorizados. Únicamente se realiza el proceso para los cinco modelos con el mejor desempeño.

*Tabla 29. Estadísticas generales de ejecución de pronósticos de ventas por cliente*

<i>Concepto</i>	<i>Descripción</i>
Tipo	Cliente
Número de Conjunto de Datos (Set)	103
Número de Modelos Calibrados	5
Tiempo Calibración (Segundos) por set	36,827
Tiempo Carga de Modelos (Segundos)	0,213
Tiempo Calculo Pronóstico (Segundos) / set	1,827
Tiempo Total (Segundos) por set	38,654
Tiempo Total (Segundos)	3981,383
Tiempo Total (Minutos)	66,356
Tiempo Total (Horas)	1,1059

### **Ventas por producto**

La Tabla 30, nos muestra un resumen de los tiempos de calibración, carga de modelos y ejecución de pronósticos para el set de datos de ventas para los productos de la categoría A. Únicamente se realiza el proceso para los cinco modelos con el mejor desempeño.

*Tabla 30. Estadísticas generales de ejecución de pronósticos de ventas por producto*

<i>Concepto</i>	<i>Descripción</i>
Tipo	Productos Categoría A
Número de Conjunto de Datos (Set)	522
Número de Modelos Calibrados	5
Tiempo Calibración (Segundos) por set	39,086
Tiempo Carga de Modelos (Segundos)	0,296
Tiempo Calculo Pronóstico (Segundos) / set	1,686
Tiempo Total (Segundos) por set	40,772
Tiempo Total (Segundos)	21282,879
Tiempo Total (Minutos)	354,71466
Tiempo Total (Horas)	5,9119

## Capítulo 14

### Conclusiones

#### Técnicas

Dentro de la perspectiva técnica, se establece:

- A excepción de los modelos ARIMA con análisis de peridiograma y ARIMA con transformación logarítmica y análisis de peridiograma, los cuales toman en promedio 50 minutos en la calibración, en general la calibración de modelos no excede los 40 segundos.
- Los modelos AR y ARIMA con menores grados de libertad (parámetros) presentaron mejores resultados en la validación del desempeño en la fase de calibración. Los pronósticos realizados con estos modelos no presentan buenos resultados en pronósticos periodos futuros (producción).
- Para los modelos Elman y Jordan, se requieren procesos adicionales, tales como la preparación de los datos (normalización) y la des normalización de datos para la evaluación del desempeño del modelo.
- Para modelos de Elman y Jordan, es necesario reestimar el modelo para cada nueva observación (Periodo a pronosticar).
- A pesar del número de muestras del conjunto de datos de entrenamiento, las redes neuronales hacen un buen acercamiento para las 73 Observaciones (Información de Ventas mensual desde enero de 2012 a diciembre de 2017).
- En términos generales, se podría mejorar el rendimiento de los modelos, a través de una calibración más detallada de los parámetros.

- Los modelos basados en redes neuronales mediante el uso de árboles de decisión (Extreme Gradient Boosting) en general tuvieron un buen desempeño. Para el procesamiento masivo de series de tiempo en la generación de pronósticos de ventas, dichas redes neuronales identificaron de manera automática los componentes de la serie temporal tales como la tendencia, estacionalidad y ciclicidad.
- La información de históricos de pronósticos de ventas, podrían ser integrada en la calibración de los modelos (Análisis multivariable), con el objetivo de mejorar la precisión de estos.

## **Negocio**

Desde la perspectiva de negocio tenemos:

- Sensibilizar a la alta Dirección acerca de la importancia de:
  - El dato como activo estratégico para la organización
  - Los beneficios de la integración de la analítica de datos en los procesos clave de la cadena de valor
  - Integrar la analítica de datos como un componente mas de los procesos
  - La analítica de datos como estrategia de mejora de los procesos internos
- Diagnosticar los procesos operativos relacionados con pronósticos de la demanda con el fin de automatizarlos, hacerlos más eficientes y funcionales.

- Hacer un diagnóstico de la cultura organizacional incluyendo a las personas, con el fin de maximizar los beneficios cuantificables de la adopción del modelo para la organización.
- La tecnología como habilitador, es fundamental entender el modelo de negocio inicialmente y luego definir la tecnología que habilita la implementación de la estrategia.
- El desarrollo de la analítica predictiva (estrategia de datos) es fundamental en el proceso de transformación digital la cual afrontan las Organizaciones.

## **Capítulo 15**

### **Trabajos futuros**

#### **Perspectiva de negocio**

#### **Integración de los pronósticos en los procesos de gestión**

Luego de la definición del modelo de pronósticos de series de tiempo, en la mejora del proceso de definición del plan y presupuesto de ventas en la organización, y de la definición de la estrategia para la implementación, se propone la implementación de la fase II y III, de acuerdo con el alcance general del proyecto. Se sugiere iniciar con la integración de los procesos de compras y gestión de inventarios en los pronósticos de ventas y la planeación de la demanda. La integración de estos procesos y la información relacionada implica el análisis no solo del importe y cantidades, sino restricciones a nivel de tiempos de reposición, precio de compra, ciclo de vida de producto, costos de logística, proceso de importación, y demás, lo cual implica un análisis multivariable. Los enfoques evaluados de redes neuronales permitirían dar solución a este nuevo planteamiento. Identificar nuevos procesos de negocio que podrían hacer uso de la solución tales como, horas por centro de trabajo (Producción), demanda de personal, (Gestión Humana), mantenimiento predictivo (Mantenimiento), etc.



Figura 98. Trabajos futuros desde la perspectiva de negocio. Elaboración propia.

## Perspectiva tecnológica

### Administración de modelos predictivos

Luego de la entrega de la administración general de la solución al área de informática, se debe asegurar que no se presente sobreajustes a los modelos calibrados en pronósticos posteriores (plan de calibración de modelos). Se debe asegurar la inclusión de información de periodos adicionales de ventas (Datos de ventas posteriores a periodo de corte para el análisis y definición de modelos) en el proceso de ajuste de modelos. Se debe asegurar la medición del desempeño del modelo para cada uno de los frentes establecidos. La información de pronósticos y valor final por juicio de experto en la consolidación del plan de ventas, podrían ser incluidos en el análisis multivariable propuesto. El proceso debería ser documentado de acuerdo con los estándares definidos en el sistema de gestión de calidad de la organización. Se debería definir y documentar el modelo de servicio de acuerdo con el estándar ITIL, relacionados con el proceso y los activos de información identificados.



Figura 99. Trabajos futuros desde la perspectiva del área de tecnología. Elaboración propia.

## Lista de referencias

- Azevedo, Ana; Santos, Manuel. (2008). KDD, SEMMA and CRISP-DM: a parallel overview. IADIS European Conference Data Mining, enero, 182-185.
- Aburto L., Weber R.; (2007). Improved supply chain management based on hybrid demand forecasts. *Applied Soft Computing*, 7, 136–144.
- Arango Marín, J., Giraldo García, J., Castrillón Gómez, O.; (2013). Gestión de compras e inventarios a partir de pronósticos Holt-Winters y diferenciación de nivel de servicio por clasificación ABC. *Scientia Et Technica*, 4, 743-749.
- Chapman, S. (2006). *Planificación y control de la producción*. 1st ed. México: Pearson Educación.
- Carbonneau R., Laframboise K., Vahidov R.; (2008). Application of machine learning techniques for supply chain demand forecasting. *European Journal of Operational Research*, 184, 1140–1154.
- Cuatrecasas Arbós, L. (2000). *Gestión de la calidad total*. 1st ed. Sunny Isles Beach: Ediciones Díaz de Santos: 250- 251.
- Curto Díaz, J. and Conesa Caralt, J. (2012). *Introducción al business intelligence*. Barcelona: Editorial UOC.
- Du, W., Xu, S., Kit, C.; (2014). Time series forecasting by neural networks: A knee point-based multiobjective evolutionary algorithm approach. *Journal of Expert Systems with Applications*, 41, 8049–8061.
- Etxeberria, J. and Tejedor Tejedor, F. (2005). *Análisis descriptivo de datos de educación*. 1st ed. Madrid: La Muralla, Pág. 32.
- Guazzelli Alex. (2012). *Predicciones sobre el futuro*. Recuperado de <https://www.ibm.com/developerworks/ssa/industry/library/ba-predictive-analytics1/>. Consultado, 2 de febrero de 2017.
- Kast, F. and Rosenzweig, J. (1987). *Administración en las organizaciones: enfoque de sistemas y de contingencias*. 1st ed. México: McGraw-Hill Interamericana: 217.
- Krajewski, L., Ritzman, L., Malhotra, M. and Krajewski, L. (2008). *Administración de operaciones: estrategia y análisis*. 5ta. ed. México: Pearson Educación: 521-523.
- Murray, P., Agard, B. and Barajas, A.; (2015). Forecasting Supply Chain Demand by Clustering Customers. *IFAC-PapersOnLine*, Abril, 1834-1839.

- Nigel Da Costa Lewis (2017). *Neural Networks for Time Series Forecasting with R*, CreateSpace Independent Publishing Platform,.
- Paul Gerrard, Radia M. Johnson. (2015). *Mastering Scientific Computing with R*, Packt Publishing.
- Putler, D. S. and Krider, R. E. (2012). *Customer and Business Analytics: Applied Data Mining for Business Decision Making Using R*. 1st ed. Boca Raton, FL: CRC Press.
- Robert H. Shumway, David S. Stoffer, (2017). *Time Series Analysis and Its Applications*, Springer, 201.
- Tianqi Chen and Carlos Guestrin (2016). XGBoost: A Scalable Tree Boosting System. In 22nd SIGKDD Conference on Knowledge Discovery and Data Mining
- Vermorel Joannès. (2013). Precisión de pronóstico (definición e ideas claves). Disponible en: <https://www.lokad.com/es/definicion-precision-de-pronostico>. Consultado, 4 de marzo de 201.
- Vermorel Joannès. (2012). The best forecast error metric. Disponible en: <http://blog.lokad.com/journal/2012/11/19/the-best-forecast-error-metric.html>. Consultado, 6 de marzo de 2017.
- Wang, W., Xu, Q; (2014). Bayesian combination forecasting model: A Bayesian Combination Forecasting Model for Retail Supply Chain Coordination. *Journal of Applied Research and Technology*, abril, 315-324.
- Yu-Wei, Chiu David (2015). *Machine Learning with R Cookbook*, Packt Publishing.

## Apéndice